ARTICLES

EMBEDDED OPTIONS AND THE CASE AGAINST COMPENSATION IN CONTRACT LAW

*Robert E. Scott**
*George G. Triantis***

Although compensation is the governing principle in contract law remedies, it has tenuous historical, economic, and empirical support. A promisor's right to breach and pay damages is only a subset of a larger family of termination rights that do not purport to compensate the promisee for losses suffered when the promisor walks away from the contemplated exchange. These termination rights can be characterized as embedded options that serve important risk management functions. We show that sellers often sell insurance to their buyers in the form of these embedded call options, and that termination fees, including damages, are in essence option prices. Furthermore, we explain why compensation is of little relevance to the option price agreed to by the parties, which is a function of the option's value to the buyer, its cost to the seller, and the market in which they transact. We propose, therefore, a novel justification for why penalty liquidated damages may be higher than the seller's costs: They are option prices that reflect the value of the options to the buyer. The regulation of liquidated damages is thus tantamount to price regulation—a function outside the realm of contract law. Moreover, in light of the heterogeneity among optimal option prices, this Article also makes the case against the expectation damages default rule. In thick markets, we argue for enforcing the parties' risk allocation with market damages. In thin markets, we propose the default rule should encourage parties to agree explicitly to termination rights, including breach damages, by the threat of specific performance of their contemplated exchange or, in the case of consumers, by a default rule that provides them a termination option at no cost.

INTRODUCTION

Compensation is the governing principle in contract law remedies. This principle shapes the key doctrines that specify the consequences of breach. Expectation, the default measure of damages that derives from the compensation principle, aims to put the promisee in the position she

* David and Mary Harrison Distinguished Professor, University of Virginia School of Law; Justin W. D'Atri Visiting Professor of Law, Business and Society, Columbia Law School.

** Perre Bowen Professor, University of Virginia School of Law.

We are grateful to workshop participants at Columbia, George Mason, Southern California, Virginia, and Yale law schools, and particularly Barry Adler, Ian Ayres, Victor Goldberg, Edward Iacobucci, Bruce Johnson, Eric Posner, Alan Schwartz, Paul Stephan, and Eric Talley for helpful comments on earlier drafts. We also thank Jason Bregge and Maruti Racherla for valuable research assistance.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

would have occupied had the promisor performed. Alternatively, specific performance is available at the option of the promisee only when the court believes money damages are inadequate compensation for her loss. Although parties may agree to liquidated damages, contract doctrine instructs them to abide by the compensation norm. But despite its profound influence on contemporary contract law doctrine, the compensation principle has tenuous historical, economic, and empirical support. Its evolution in the common law resulted primarily from ill-conceived path dependence; compensation is virtually ignored in the theoretical analysis of efficient contract design; and compensation plays little role in the contracts actually negotiated by commercial parties and agreed to by consumers. As a result of an unfortunate turn in history, lawmakers view contract damages as compensation for wrongs. This has impeded both the efficient evolution of default remedies and the efficient regulation of liquidated damages.

Rather than conceiving of damages as compensation, the right to breach and pay damages is better understood as a valuable option sold by the promisee to the promisor. Indeed, the right to breach is only a subset of a broader category of termination rights that gives one party an option to walk away from the contemplated exchange.[1] A firm offer or unilateral promise, for example, grants the promisee such an option. Broad warranties, such as satisfaction-or-your-money-back provisions, give buyers similar options. Requirements, output, or installment contracts grant one party substantial discretion to determine the contract quantity. And

---

1. It is well known that contract damages effectively give the promisor an option between performing the promise or breaching and paying damages. The classic statement is by Justice Holmes:

> Nowhere is the confusion between legal and moral ideas more manifest than in the law of contract. Among other things, here again the so called primary rights and duties are invested with a mystic significance beyond what can be assigned and explained. The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it,—and nothing else.

Oliver Wendell Holmes, The Path of the Law, 10 Harv. L. Rev. 457, 462 (1897).

Over the past decade, legal scholars have begun to analyze contract remedies explicitly in terms of options. In this article, we extend prior analyses to consider the full range of contractual options to terminate. Our normative claims follow from this broader perspective. For prior work on contract options, see generally Victor P. Goldberg, Discretion in Long-Term Open Quantity Contracts: Reining in Good Faith, 35 U.C. Davis L. Rev. 319 (2002) [hereinafter Goldberg, Discretion]; Victor P. Goldberg, The Net Profits Puzzle, 97 Colum. L. Rev. 524 (1997) [hereinafter Goldberg, Net Profits Puzzle]; Avery Wiener Katz, The Efficient Design of Option Contracts: Principles and Applications, 90 Va. L. Rev. (forthcoming Dec. 2004) (manuscript on file with the *Columbia Law Review*); Paul G. Mahoney, Contract Remedies and Options Pricing, 24 J. Legal Stud. 139 (1995); Robert E. Scott, The Case for Market Damages: Revisiting the Lost Profits Puzzle, 57 U. Chi. L. Rev. 1155 (1990) [hereinafter Scott, Market Damages]; Alexander J. Triantis & George G. Triantis, Timing Problems in Contract Breach Decisions, 41 J.L. & Econ. 163 (1998); George G. Triantis, The Effects of Insolvency and Bankruptcy on Contract Performance and Adjustment, 43 U. Toronto L.J. 679 (1993) [hereinafter Triantis, Insolvency and Bankruptcy].

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

a contract may provide that one party has the right to terminate, cancel, renew, return, or redeem goods.

Either or both parties to a contract, therefore, commonly enjoy the right to terminate at some cost. For the purposes of analysis and argument, we focus on the option held by a buyer of goods or services.[2] The buyer holding an option has the right to avoid the exchange by paying either a termination fee or damages. The price of an embedded option is determined just as the price of any other product: It is a function of the option's value to the option holder, the cost to the option writer, and the competitiveness of the market in which they transact. Options are essentially insurance contracts that divide risks according to their respective exercise prices.[3] For any given exercise price, the option price divides between the parties the surplus created by the option—namely, the difference between the option's value to the buyer and its cost to the seller. The parties should choose a pairing of option price and exercise price that maximizes this surplus. Sometimes the option price is fixed by the contract (e.g., in the form of a nonrefundable deposit or termination fee), and at other times it is left to be judicially determined as damages for breach of contract. Given the great variety of conditions under which parties contract for this option, it should not be surprising that commercial and consumer contracts contain a wide range of option prices. We explain the heterogeneity in option prices and argue that they are rarely equivalent to the measure of the seller's expectation in a completed sale.

Consider an electronics store that sells television sets for $400 and offers full refunds for any returns made within thirty days. This contract gives the buyer a free thirty-day option to purchase the television set for $400. A buyer who does not know whether her family needs a television set or will like this particular model values the opportunity to return it free of charge. Thus, the option is valuable to her because she is uncertain as to the value of the television set to her family. The retailer in this case bears significant costs in accepting returns, including the cost of re-

2. Several previous articles characterize the right to breach as the buyer holding a put option on the agreed-upon exchange that she can exercise by paying damages to the seller. By exercising the put, the buyer in effect sells the contract good or service back to the seller for the assessed damages liability. E.g., Mahoney, supra note 1, at 140; Triantis & Triantis, supra note 1, at 168–69 & n.15; Triantis, Insolvency and Bankruptcy, supra note 1, at 680–84. In this Article, however, we analyze the buyer as effectively incurring an obligation to pay damages in exchange for a call option on the subject matter of the contract. The buyer's prospective liability for damages is effectively the price of the call option, as if the buyer makes a nonrefundable deposit or payment for the call option and pays an additional price to exercise it. The exercise price of the option is the difference between the contract price and the damages liability. For discussion, see infra text accompanying notes 113–115.

3. In his Nobel lecture, Robert Merton stated: "When [an option is] purchased in conjunction with ownership of the underlying asset, it is functionally equivalent to an insurance policy that protects its owner against economic loss from a decline in the asset's value below the exercise price for any reason . . . ." Robert C. Merton, Applications of Option-Pricing Theory: Twenty-Five Years Later, 88 Am. Econ. Rev. 323, 336–37 (1998).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

ceiving, inspecting, and reselling the returned goods, often through a discount outlet or internet sale. Yet, the retailer often does not charge for the option.[4] Such "free" options are particularly interesting for our purposes because these contracts make no attempt to compensate the seller for losses it suffers when the buyer walks away from the contemplated exchange.[5]

Under many other termination provisions, buyers pay positive option prices in the form of fees or damages. These fees sometimes exceed the compensatory amount—for example, the seller's foregone profit. Consumers are familiar with instances of such overcompensation in a variety of transactions. For example, economy fares on airlines are typically conditioned on a penalty of $100 if the passenger chooses to cancel and apply the ticket price against the fare of another flight. The penalty applies even on flights that are overbooked and almost certain to be full. Thus, the penalty is not simply compensation for the airlines' losses.

Indeed, casual observation reveals that termination fees in both commercial and consumer contracts regularly depart from the compensatory amount in both directions. Retailers often have the right to return unsold merchandise to the wholesaler or distributor. Rights of return are common, for example, in the retailing of books, journals, newspapers, musical compact discs, jewelry, and cigarettes.[6] Aircraft manufacturers permit purchasers to cancel orders or to change the type of aircraft at no charge even after the manufacturer has made a significant investment in production.[7]

In this Article, we offer an explanation for these contracting patterns based on the insurance (or risk management) function of embedded op-

---

4. Another example of a free option is the right of consumer borrowers to prepay their loans or mortgages without compensating their lenders for consequential losses that might be caused, for example, by declines in the market rate of interest. Similarly, a borrower who defaults is liable to pay the accelerated balance owing, but not compensation for the lender's foregone opportunities when market rates have dropped since the loan. We briefly consider the motivation behind the granting of free options at infra notes 119, 151 and accompanying text.

5. It is inaccurate to conclude that the cost of the option is included in the overall contract price because the buyer does not pay this price if she walks away from the option. Where the seller does not charge the buyer for an unexercised option, the seller recovers the cost from buyers who exercise the option. This cross-subsidization leads to adverse selection and moral hazard issues discussed later in the Article. In many of these cases, the option is in fact not free because the seller holds the buyer's payment of the contract price during the option term and, if the buyer has possession of the good during that period, she often must also bear the cost of bringing the product back to the seller. See infra Part II.D.

6. Eugene Kandel, The Right to Return, 39 J.L. & Econ. 329, 330 (1996).

7. John Stonier, The Change Process, *in* Real Options: A Practitioner's Guide 28, 47 (Tom Copeland & Vladimir Antikarov eds., 2001); John Stonier & Alexander J. Triantis, Natural and Contractual Real Options: The Case of Aircraft Delivery Options, *in* Real Options Applications: Proceedings of the First Milan International Workshop on Real Options 159, 160 (Alberto Micalizzi & Lenos Trigeorgis eds., 1999).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

tions.[8] It is well known that compensatory remedies insure the buyer against the seller's breach and thereby against the risk of fluctuations in the seller's *cost* of performance that lead it to breach. The buyer, however, still bears the risk of fluctuations in the *value* of the seller's performance. Options created by termination rights insure buyers against this risk and thereby may promote the risk management objectives of business contractors.

When the buyer's option comes in the form of the right to breach and pay expectation damages, the buyer's liability on termination is a function of the profit that the seller would gain from the exchange. Under an expectation damages rule, therefore, the buyer shrugs off some of the risk of fluctuations in performance value, but assumes some of the risk in fluctuations in the seller's profit. If the buyer contracts for the right to terminate and pay a fixed amount rather than expectation damages, the buyer can also avoid the risk associated with the seller's cost. The buyer may be prepared to pay a premium to be able to shift both the risk in the value of performance and the risk in the seller's costs. In these cases, the parties may agree to pay fixed liquidated damages that are greater than the (ex ante) expected amount of the seller's loss from breach. Under the current penalty rule, however, courts will refuse to enforce this term.[9] Our analysis suggests a benign explanation for this supercompensatory damages term: It is the product of a negotiated sale of insurance from the seller to the buyer in the form of an embedded option.

Given the close link between options and insurance, it should not be surprising that the optimal terms of embedded options are a function of considerations that determine insurance contracts—namely, risk-bearing capacity, adverse selection, and moral hazard. Although we discuss the role of these considerations in the structuring of embedded options, our principal contribution is to show that many contracts contain embedded options and that the optimal terms and prices of these options are heterogeneous. This analysis is consistent with our observation of a wide variety of termination provisions in practice.

Our analysis has important normative implications for the default rules of contract damages and for the freedom of parties to contract away from the defaults. Termination provisions serve valuable risk management objectives when they depart from the compensation principle. The

8. In this Article we use the terms "insurance" and "risk management" interchangeably. Our focus is on the role of embedded options in facilitating parties' ability to hedge risks in a wide variety of contracts. We do not assume that this hedging is a response to risk aversion. Rather, we argue that embedded options are ways of managing risks in incomplete markets that are attractive to parties regardless of their risk profiles. See infra Part II.B.

9. The penalty rule voids any unreasonably large liquidated damages term. Reasonableness, in turn, is determined by the anticipated or actual harm caused by the breach and the difficulty of proving actual loss. See infra text accompanying notes 20–23.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

characterization of breach damages as the price of an option that yields value to the buyer reinforces the criticisms of the penalty rule that have been raised by contracts scholars.[10] The recent litigation concerning late fees charged by Blockbuster video stores is illustrative.[11] There, the plaintiffs argued that the fees were unenforceable penalties.[12] But the court recognized that the right to extend the rental period was valuable, that a rational consumer would pay a price for this right that might be higher than the cost to Blockbuster, and that courts should not engage in regulation of this price.[13] We argue that courts similarly should enforce all termination provisions, regardless of their form.

Courts are not well suited to set default contract damages either. It is unlikely that any given damages default will reflect the option price that most parties would adopt in their contracts. In most contract relationships, an expectancy default rule is no more appropriate with respect to damages than with respect to the price of any other good, service, or contract term for which there is no established market price. We suggest, moreover, that the particular salience of a damages default—particularly one founded on an entrenched compensation principle—discourages parties from writing explicit termination provisions and deters the courts from specifically enforcing them. Thus, we argue in favor of a bargain-facilitating, instead of a majoritarian, approach. Where commercial parties have failed to provide expressly for termination rights, we propose that the courts specifically enforce the exchange. In the context of consumer transactions, where merchant sellers typically draft the contracts, the default should give the consumer a free option if the parties fail to make an express provision regarding termination rights.

This Article proceeds as follows. In Part I, we contrast the contemporary dominance of the compensation principle of contract damages against its tenuous historical roots, its marginal relevance in contract economics, and its limited use in commercial and consumer contracts. Part II characterizes the termination provisions available to a buyer—including the right to breach and pay damages—as call options that serve an important risk management function. We demonstrate that option

10. For a discussion of the scholarly criticism of the penalty rule, see infra note 103. The compensation principle is also responsible for misleading courts as well as scholars in various doctrinal debates concerning alternative measures of expectation damages. See infra text accompanying notes 168–172.

11. Pickens v. Blockbuster, Inc., No. A102626, 2004 WL 339594 (Cal. Ct. App. Feb. 24, 2004).

12. Id. at *1.

13. The California Court of Appeal held that the test for whether the late fees were liquidated damages or a provision for alternative performance is whether the customer would have a "realistic and rational choice" between two alternative performances (returning the video when due or keeping it). Id. at *3. The court then ruled that "[t]his is not a case in which no rational person would choose to pay the greater liability." Id. The court cited Blank v. Borden, 11 Cal. 3d 963, 971 (1974), as authority. *Blockbuster*, 2004 WL 339594, at *3.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

prices, including breach damages, might take a wide range of values. We specify factors that explain the variations in the prices paid for these options and show that these prices may appear either undercompensatory or overcompensatory to a court. Part III develops the implications for optimal default rules for breach of contract summarized above. We conclude that contracting parties should be free to set the prices for embedded options, and that the law should not provide default damages rules that effectively establish prices for the parties when they fail to do so for themselves.

I. Reconsidering the Compensation Principle in Contract Law

A. *Doctrinal Dominance of the Compensation Principle*

The doctrinal view of contract breach is that it is a wrong that is remedied by compensating the victim for her loss.[14] This view is reflected in the default provisions for damages and the constraints imposed on the freedom of parties to stipulate other measures of damages by contract. The Restatement (Second) of Contracts, for example, states that "[t]he traditional goal of the law of contract remedies has not been compulsion of the promisor to perform his promise but compensation of the promisee for the loss resulting from breach."[15] Since the classic article by Fuller and Perdue in 1936,[16] courts and scholars typically have considered three alternative bases for determining the promisee's entitlement to compensation: restitution, reliance, and expectation. The dominant measure is expectation damages—the payment necessary to put the promisee in as good a position as if the promisor had performed.[17] Over the past 150 years, the courts have expanded the availability of compensation for the promisee's consequential losses, subject only to the relatively modest constraints of mitigation, foreseeability, and uncertainty.[18] Spe-

14. See, e.g., U.C.C. § 1-305(a) (2003) (formerly § 1-106) ("The remedies provided by [the Uniform Commercial Code] must be liberally administered to the end that the aggrieved party may be put in as good a position as if the other party had fully performed . . . ."); Restatement (Second) of Contracts § 347 & cmt. a (1981) ("Contract damages are ordinarily based on the injured party's expectation interest and are intended to . . . put him in as good a position as he would have been in had the contract been performed."). Although the compensation principle is firmly enshrined in the Code and the Restatement, we suggest in the next section that contract law embraced the compensation principle relatively late in its common law development.

15. Restatement (Second) of Contracts ch. 16 introductory note at 100; see also U.C.C. § 1-305 cmt. 1 (explaining common view that central purpose of contract damages is compensation); E. Allan Farnsworth, Legal Remedies for Breach of Contract, 70 Colum. L. Rev. 1145, 1147 (1970) (same).

16. L. L. Fuller & William R. Perdue, Jr., The Reliance Interest in Contract Damages: 1, 46 Yale L.J. 52 (1936).

17. Restatement (Second) of Contracts §§ 344, 347(a). "The initial assumption is that the injured party is entitled to full compensation for his actual loss." Id. ch. 16 topic 2 introductory note at 109.

18. In a recent article, Richard Craswell observes that courts appear willing to soften or disregard these traditional limitations when they find that the breach was willful.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

cific performance is available only when the court determines that money damages are inadequate to compensate for the promisee's loss.[19]

The most notable characteristic of the compensation principle is its expansive reach. The compensation default applies in all contexts and, in addition, purports to justify the legal regulation of parties' ability to stipulate damages expressly in their contract. Although most of contract law provides default rules from which parties are free to contract away, remedial defaults carry heavier presumptive weight than other provisions. Contract breach terms may not deviate from the compensation principle. According to doctrinal statements, the only permissible ground for stipulating damages is the anticipated difficulty of measuring the promisee's loss. For instance, the Restatement and Article 2 of the Uniform Commercial Code each provide that stipulated damages must be an amount that is reasonable in light of the anticipated or actual loss caused by the breach and the difficulties of proving the amount of loss.[20] In practice, the courts strike down supercompensatory liquidated damages far more often than they police undercompensatory limitations on damages.[21] In

---

Richard Craswell, Against Fuller and Perdue, 67 U. Chi. L. Rev. 99, 138–43 (2000) [hereinafter Craswell, Against Fuller and Perdue]. Craswell cites Arthur Corbin as noting "'a lesser degree of certainty will be required as against one whose breach is described as 'willful' or is motivated by malice or avarice than against one whose breach was due to misfortune and whose efforts to perform were honest and in good faith.'" Id. at 140 (quoting 5 Arthur Linton Corbin, Corbin on Contracts § 1020 (1964)). Craswell also speculates that courts may choose reliance damages as a means of effecting some sharing of losses between contracting partners (for example, in response to an unforeseen contingency). Craswell, Against Fuller and Perdue, supra, at 150.

19. Restatement (Second) of Contracts § 359; see also Robert E. Scott & Jody S. Kraus, Contract Law and Theory 115–17, 988–95 (rev. 3d ed. 2003) (discussing when specific performance is appropriate).

20. U.C.C. § 2-718(1) ("Damages for breach by either party may be liquidated in the agreement but only at an amount which is reasonable in the light of the anticipated or actual harm caused by the breach, the difficulties of proof of loss . . . ."); Restatement (Second) of Contracts § 356 cmt. a ("The parties to a contract may effectively provide in advance the damages that are to be payable in the event of breach as long as the provision does not disregard the principle of compensation."). The modern penalty rule found in the Code and the Restatement is derived from a line of common law cases invalidating any stipulated damages where the amount specified exceeded the "just compensation for the loss or injury actually sustained." Jaquith v. Hudson, 5 Mich. 123, 133 (1858). For discussion, see Charles J. Goetz & Robert E. Scott, Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach, 77 Colum. L. Rev. 554, 555–57 (1977) [hereinafter Goetz & Scott, Liquidated Damages] and infra text accompanying notes 70–72.

21. See 5 Corbin, supra note 18, § 1068 ("[W]ith certain exceptions, the courts see no harm in express agreements limiting the damages to be recovered for breach of contract. Public policy may forbid the enforcement of penalties against a defendant; but it does not forbid the enforcement of a limitation in his favor." (footnote omitted)); id. § 1068 (Supp. 1998) (citing cases in support); William F. Fritz, "Underliquidated" Damages as Limitations of Liability, 33 Tex. L. Rev. 196, 203–06 (1954) (same). Our own review of reported cases reveals that the incidence with which courts strike down penalties far exceeds their rejection of undercompensatory damages provisions. See, e.g., Martin Rispens & Sons v. Hall Farms, 621 N.E.2d 1078, 1087 (Ind. 1993) ("[C]ourts have rejected claims that

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

cases in which it is especially difficult to measure the promisee's loss, a court may order specific performance or an injunction even in the face of a contractual provision for liquidated damages.[22] This additional basis for setting aside liquidated damages provides further evidence that the law regards measurement difficulty as the sole legitimate reason for deviating from the default of expectation damages. Contracting parties sometimes attempt to circumvent the regulation of liquidated damages by providing explicitly for alternative contracts, such as take-or-pay and pay-or-play provisions, that purport to give the promisor a choice between performance and a monetary alternative. But even these drafting techniques are more likely to succeed when the alternative payment has a compensatory basis, such as reimbursing fixed costs incurred by the seller.[23]

The universality of the compensation principle has largely shielded the penalty doctrine and the expectation damages default from serious scrutiny by legal scholars. But despite its doctrinal dominance, the compensation principle rests on tenuous foundations. As we argue below in the remainder of Part I, historical analysis, economic theory, and commercial practice all suggest that the deference accorded to the notion of compensation for breach is unjustified. Moreover, our analysis of the economic function of embedded options that follows in Part II shows that the compensation principle is inconsistent with many of the bargains that contracting parties would otherwise choose to make.

B. *Historical Roots of Compensation in Breach Remedies*

Although now firmly entrenched in doctrine, the contemporary understanding of the compensation principle is a recent development in contract law. Well into the nineteenth century, contract enforcement was heterogeneous. The common law courts provided separate remedies for discrete contracting categories. Thus, for example, courts awarded reimbursement damages for breach of informal contracts in thin markets; awarded market damages in thick-market contexts, such as commodities or stock transactions; specifically enforced contracts under seal; and gen-

---

contractual limitations of remedy are substantively unconscionable."); cases cited infra note 109.

22. Restatement (Second) of Contracts § 361 cmt. a.

> Merely by providing for liquidated damages, the parties are not taken to have fixed a price to be paid for the privilege not to perform. *The same uncertainty as to the loss caused that argues for the enforceability of the provision may also argue for the inadequacy of the remedy that it provides.*

Id. (emphasis added).

23. Compare Prenalta Corp. v. Colo. Interstate Gas Co., 944 F.2d 677, 690 (10th Cir. 1991) (upholding take-or-pay clause as enforceable alternative contract), with Lake River Corp. v. Carborundum Co., 769 F.2d 1284, 1292–93 (7th Cir. 1985) (finding minimum guarantee clause void as a penalty).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

erally enforced parties' intentions regarding liquidated damages.[24] Beginning in the nineteenth century, however, the courts began to collapse these categories under an expanding compensation principle that they adopted from tort law. Thereafter, the compensation notion gradually mutated to include full expectancy damages, and this broad conception of compensation spread to each of the previously discrete transactions. This development severely narrowed the freedom of contracting parties to stipulate damages and contract for specific performance.

1. *The Roots of the Compensation Principle.* — At early common law, there was no cause of action for breach of an informal (unsealed) executory promise. The only actions available for breach of contract were the action for debt and the action in covenant (for promises under seal).[25] The notion of compensation was foreign to either action. The action for debt lay only for the recovery of a sum certain. The plaintiff sought relief for a debt that was due and owing, fixed by the parties' prior agreement, and the court in no sense awarded compensation for breach of contract.[26] But the evolution of commercial exchange during the late middle ages led the English common law courts eventually to recognize a promisee's right to recover for breach of an informal promise by bringing an action in assumpsit.[27]

Assumpsit had developed initially to provide a cause of action for the negligence of a bailee or carrier for hire.[28] The principle of compensa-

---

24. Throughout this Article, we continue to draw the distinction between thin-market contracts (where one or both parties make a relation-specific investment) and thick-market contracts (where risk allocation rather than investment is the primary objective). See infra note 139.

25. James Barr Ames, Debt, *in* Lectures on Legal History 88, 92 (1913) [hereinafter Ames, Debt]; James Barr Ames, Simple Contracts Prior to Assumpsit, *in* Lectures on Legal History, supra, at 122, 122–23; A.W.B. Simpson, A History of the Common Law of Contract 47 (1987) [hereinafter Simpson, History].

26. Ames, Debt, supra note 25, at 88–89. Where a seller tendered goods to a buyer and the buyer refused to accept delivery, the seller was able to sue in debt for the purchase price and force the buyer to take delivery of the goods (for which title had passed under the contract). Alternatively, if the buyer tendered the purchase price and the seller refused to transfer goods that were then available, the buyer's only recourse was to bring an action in equity for specific performance since the remedy at law was inadequate.

27. Thomas F.T. Plucknett, A Concise History of the Common Law 637–46 (5th ed. 1956). In the absence of legal enforcement, parties relied on the law merchant and medieval trade fairs as a means of self-enforcing informal promises. Avner Grief, Informal Contract Enforcement: Lessons from Medieval Trade, *in* 2 The New Palgrave Dictionary of Economics and the Law 287, 288–89 (Peter Newman ed., 1998) [hereinafter New Palgrave Dictionary].

28. Simpson, History, supra note 25, at 210–15. The traditional tort action of trespass on the case would not permit recovery for negligent bailment because the plaintiff might have been equally careless in entrusting a third party with his property. Id. The action in assumpsit for negligent bailment was first recognized in the Humber Ferryman Case, 22 Lib. Ass., pl.41 (1348), reprinted in Simpson, History, supra note 25, at 211. The plaintiff alleged that defendant bailee *undertook* to carry his goods safely. Id. The failure to perform this undertaking was the gravamen of the action, and, as was traditional in tort actions, the resulting injury to the plaintiff's property required *compensation*. Id. at 211–12.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

tion that supported the action in assumpsit was thus a tort notion premised on the idea of ex post redress for a harm committed by the defendant.[29] Over time, the action in assumpsit was extended to nonperformance of certain promissory undertakings.[30] Specifically, the action in assumpsit for breach of promise lay for plaintiffs who had either conferred benefits or had taken action in preparation for performance in reliance on the defendant's promise. In either case, a plaintiff who was seeking relief via assumpsit for breach of an informal promise was asking for compensation under a theory of *reimbursement* for the loss of that which had been given (directly or indirectly) to the promisor.[31]

For many years, the orthodox view among legal historians was that the final stage of the development of assumpsit—the enforcement of purely executory promises—matured in England sometime in the sixteenth or seventeenth centuries.[32] But the weight of evidence suggests that the enforcement of executory contracts occurred much later, at the end of the eighteenth century (in England) and the beginning of the nineteenth century (in the United States).[33] Throughout the eighteenth century, contract law was still dominated by the action in debt: Exchange

---

29. James Barr Ames, Express Assumpsit, *in* Lectures on Legal History, supra note 25, at 129, 130. A parallel line of cases permitted recovery in deceit for a false warranty for goods sold and delivered. This action was also, in its origin, a pure action in tort. Id. at 136–37.

30. Initially, a plaintiff was allowed to bring assumpsit only where the defendant performed his promise unskillfully (e.g., a carpenter who undertook to build a house for the plaintiff and performed poorly). Id. at 139–41. Subsequently, the English courts held that a plaintiff could recover in assumpsit for the promisor's failure to act altogether. Id. at 142.

31. During this early period of the action in assumpsit, a plaintiff could bring an action for breach of promise independent of the doctrine of consideration and the concept of exchange. Id. at 130. The early notion of *special assumpsit* (the contract action) did not require a quid pro quo as was required for an action for debt, which was explicitly tied to the notion of exchange. Id. at 147.

32. Morton Horwitz, The Historical Foundations of Modern Contract Law, 87 Harv. L. Rev. 917, 919–20 (1974); see also E. Allan Farnsworth, Contracts 18 (4th ed. 2004) [hereinafter Farnsworth, Contracts] (noting "there was a growing inclination among common law judges . . . to make a major extension in the action of assumpsit by enforcing such promises").

33. See Horwitz, supra note 32, at 922–23 (arguing that enforcement of executory promises did not occur until the rise of industrialization and the development of commercial markets in the late eighteenth and early nineteenth centuries). Horwitz's basic thesis—that prior to the industrial revolution the common law of contract was dominated by notions of equity and fairness and that it was thereafter adapted to legitimize the inequalities of the nineteenth-century market economy—has been vigorously contested. See, e.g., A. W. B. Simpson, The Horwitz Thesis and the History of Contracts, 46 U. Chi. L. Rev. 533 (1979). Simpson's critique, however, does not challenge the basic point relevant to the emergence of the compensation principle—that courts did not regularly enforce executory contracts until the nineteenth century. Rather, as we discuss below, the penal bond was the preferred device for legal enforcement of commercial exchange transactions in the seventeenth and eighteenth centuries. See infra text accompanying notes 42–52.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

was not conceived of in terms of future returns and thus expectation damages were not recognized on either side of the Atlantic.[34] During this pre-market period, courts almost universally declined to instruct juries on damages measures or to revise damages judgments (whether excessive or inadequate).[35] The common law courts that granted recovery for an action in debt were, in essence, specifically enforcing the actual bargain that the parties had struck.

Executory contracts thus were not enforced in the United States until the early nineteenth century.[36] This development coincided with a period of commercial expansion and with the emergence of markets in stock transactions and commodities.[37] Courts began awarding market-based damages for failure to deliver stock certificates in a rising market[38] and for breach of fixed-price forward contracts for the delivery of commodities.[39] This link between commodity and stock transactions and contract law was the major step in the development of a market damages default rule for breach of thick-market contracts. Contract thereafter became an instrument for managing the exogenous price changes in well-developed markets.[40]

---

34. See Horwitz, supra note 32, at 921–22. Horwitz states that only two reported eighteenth-century English cases raise the issue of expectation damages. Id. at 921. In *Fleureau v. Thornhill*, for example, the court limited the plaintiff to restitution damages, holding that the plaintiff could not be "entitled to any damages for the fancied goodness of the bargain, which he supposes he has lost." 96 Eng. Rep. 635, 1078 (C.P. 1776). In the American colonies, only a few actions for breach of executory contracts were brought before the Revolution. See, e.g., Boehm v. Engle, 1 Dall. 15, 16 (Pa. 1767) (allowing seller to sue for contract price after breach of a contract for sale of land).

35. Theodore Sedgwick, A Treatise on the Measure of Damages 200–01 (New York, J.S. Voorhies 2d ed. 1852). Chancellor Kent articulated this principle of contract damages as compensation being fixed by the jury "with a moderation agreeable to equity and good conscience, and when the claims and pretensions of each party can be duly attended to." Seymour v. Delancey, 6 Johns. Ch. 220, 232 (N.Y. Ch. 1822).

36. See Sands v. Taylor, 5 Johns. 395 (N.Y. 1810). Under the older common law rule, when a buyer breached a contract to purchase goods, the seller would have been required to tender the contract goods and sue for the contract price. Id. at 400–01. But, in *Sands*, the seller covered on the market by reselling the goods to a third party and then sought damages based upon the contract-market differential. Id. at 397. The court conceded that this was a case of first impression in the United States and granted market damages to the plaintiff. Id.

37. Horwitz, supra note 32, at 921–23.

38. See, e.g., Groves v. Graves, 1 Va. (1 Wash.) 1, 3–4 (1790).

39. See, e.g., Shepherd v. Hampton, 16 U.S. (3 Wheat.) 200, 204 (1818).

40. Horwitz, supra note 32, at 941. A market damages default rule was generalized in England in 1826 with the publication of the first treatise announcing a general rule of damages for failure to deliver goods: "In an action for [sic] assumpsit, for not delivering goods upon a given day, the measure of damages is the difference between the contract price, and that which goods of similar quality and description, bore on or about the day, when the goods ought to have been delivered." Id. at 941 n.124 (quoting Joseph Chitty, A Practical Treatise on the Law of Contracts Not Under Seal and upon the Usual Defenses to Actions Thereon 131–32 (London, S. Sweet 1826)). It was at this point that contract fully separated from property; courts, for the first time, granted promisees a property right in the contract itself.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

2. *The Penal Bond and the Roots of Judicial Oversight of Liquidated Damages.* — If legal enforcement of executory contracts was unavailable prior to the industrial revolution, how did commercial parties conduct exchange transactions? To be sure, many contracting parties relied on self-enforcement of their bargains.[41] But, in addition, commercial parties developed alternative legal mechanisms to make credible commitments. The most important of these devices was the penal bond. The bond was a sealed promise to pay a sum of money subject to an express condition subsequent.[42] The bond was made conditional on an executory promise to perform a specific service or to deliver goods or money at a time certain.[43] The advantage of this formal instrument was that the promise to pay was conclusively enforceable at its face value by an action in covenant and was an absolute obligation once the defeasing condition failed to occur.[44] In a real sense, therefore, the bond was a means of specifically enforcing a contract with the stated monetary obligation serving as security to ensure performance. According to Horwitz, at the end of the eighteenth century virtually all large business transactions took the form of two independent bonds, each of which stipulated damages for failure to perform an executory promise.[45]

Penal bonds thus effectively served as the eighteenth-century substitute for legally binding executory contracts. Indeed, at the beginning of the nineteenth century the number of bonds used to effect commercial transactions in America greatly exceeded the number of contracts that sought to enforce mutual promises.[46] The dominance of bonds, bills of exchange, and other sealed instruments meant that commercial parties had little reason to take their transactional disputes to common law courts. As a result, the modern default rule of damages for executory contracts did not develop until the mid-nineteenth century.

From the beginning of the eighteenth century, English courts adjudicating disputes over the terms of bonds had sought to distinguish between penalties—for which they would grant relief from the bond in ap-

---

41. Grief, supra note 27, at 288–89.

42. Initially, the bond developed as a debt instrument to circumvent restrictions on interest under rigid usury laws. For example, a debtor might promise under seal to pay a creditor £1,000 in six months. This promise would be subject to an express condition subsequent that it was null and void if the debtor paid the creditor £900 (the amount of the loan) on the day following the execution of the sealed promise. 2 Theophilus Parsons, The Law of Contracts 433 (Boston, Little Brown & Co. 1855); Sedgwick, supra note 35, at 392–93.

43. Sedgwick, supra note 35, at 392.

44. See A. W. B. Simpson, The Penal Bond with Conditional Defeasance, 82 L.Q. Rev. 392, 411–12 (1966) (explaining that "[t]he law governing bonds is tough law" and that such conditional bonds were almost always enforced"); see also Sedgwick, supra note 35, at 393 (noting penalty "was recoverable without any reference whatever to the actual damages incurred").

45. Horwitz, supra note 32, at 928.

46. Id. at 929.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

propriate cases[47]—and liquidated damages, which the parties were free to stipulate in the bond without interference from the courts.[48] Recall that the penalty in the bond was designed only to secure performance. It was not a sum that the promisor intended to pay. But the penalty was a poor security device because the monetary payment was independently valuable to the promisee, thus raising the risk of opportunism and induced breach.[49] In response, courts of equity policed performance under the bond to ensure that minor deviations from performance by the promisor were not used as a pretext for enforcing the face value of the bond.[50] By the end of the eighteenth century, common law courts had adopted the equity rule of relief from the bond where the amount owed greatly exceeded the loss to the plaintiff from the breach of the condition.[51] American courts followed this tradition and "chancered" these bonds.[52]

3. *The Mutation of Market Damages to Expectancy.* — During the early decades of the nineteenth century, executory contracts began to replace the bond for most transactions between commercial parties.[53] The market damages default rule that had developed in the commodities market context was adopted as well for contracts in which the parties had made

47. 6 W. S. Holdsworth, A History of English Law 663 (1924). Starting from the late seventeenth century, courts of equity began to grant claims for relief from the bond where there was only a minor breach of the stipulated conditions. Id. at 663 & n.2.

48. By the time of Lord Mansfield, English courts were predisposed to enforce liquidated damages provisions. Mansfield announced as settled the proposition that "where the covenant is 'to pay a particular liquidated sum,' a Court of Equity can not make a new covenant for a man." Lowe v. Peers, 98 Eng. Rep. 160, 162 (K.B. 1768). In 1801, Lord Eldon declared that "one cannot but lament . . . [the adoption of the principle] that if the sum would be very enormous and excessive considered as liquidated damages, it shall be taken to be a penalty though agreed to be paid in the form of contract." Astley v. Weldon, 126 Eng. Rep. 1318, 1321 (C.P. 1801).

49. See Kenneth W. Clarkson et al., Liquidated Damages v. Penalties: Sense or Nonsense?, 1978 Wis. L. Rev. 351, 368–72 (explaining induced breach). For a discussion of the hostage function of security and the attendant risk of induced breach, see Robert E. Scott, A Relational Theory of Secured Financing, 86 Colum. L. Rev. 901, 927–29 (1986).

50. See Farnsworth, Contracts, supra note 32, at 812 (stating that near the end of the seventeenth century equity principles were applied to bar award of penal sums beyond actual loss); 5 Holdsworth, supra note 47, at 293 (describing relief against penalties under the principle that it was "against conscience" that a person should recover damages in excess of incurred losses).

51. See Sedgwick, supra note 35, at 394 (describing common law courts' adoption of the equity rule of relief from the bond); Seymour D. Thompson, Penalties and Liquidated Damages, 46 Cent. L.J. 5, 6 (1898) (explaining that relief from penalties brought law into line with popular notion of unconscionability).

52. See Sedgwick, supra note 35, at 394 (describing New York's adoption of relief statutes).

53. Liquidated damages provisions in formal bonds were not well suited to contracts premised on the allocation of market risk. See Graham v. Bickham, 2 Yeates 32, 33–34 (Pa. 1795) (holding that sharp market fluctuations justify recovery for market damages in excess of the stipulation in a bond); see also Horwitz, supra note 32, at 931–32 (arguing against suitability of bonds in light of *Graham*).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

contract-specific investments. But plaintiffs seeking such "expectancy" damages for breach of contract were generally denied recovery for consequential losses.[54] Sedgwick, writing in 1852, stated that "[b]oth the English and American courts have generally concurred in denying profits as any part of the damages to be compensated for [breach of contract]."[55] Thus, he concluded that "[t]he law does not aim at complete compensation for the injury sustained; that it seeks rather to divide than to satisfy the loss, and that in cases of contract . . . the direct pecuniary damage . . . form[s] the measure of relief."[56] Justice Story, in a case involving illegal capture, explained this general principle for both tort and contract: "[A]n allowance of damages upon the basis of a calculation of profits is inadmissible. . . . The subject would be involved in utter uncertainty. . . . After all, it would be a calculation upon conjecture and not upon facts."[57]

Despite this general principle, courts during this period sometimes allowed plaintiffs to recover lost profits.[58] Sedgwick concluded, therefore, that "it is difficult to lay down any general rules in regard to the extent to which the law goes in search of resulting damages."[59] Nevertheless, he announced a general default rule governing damages for breach of contract: The breaching party is liable for losses that fairly were in contemplation of the parties at the time of contract; that is, the plaintiff must have "turned the mind of the [defendant] to the consequences likely to ensue from default."[60] The early crystallization of this rule sug-

---

54. See 2 Parsons, supra note 42, at 458–59 (collecting cases); Sedgwick, supra note 35, at 68–69 (same). Prior to the decision in Hadley v. Baxendale, 156 Eng. Rep. 145 (Ex. 1854), courts generally refused to take into consideration any damages that "remotely" resulted from the breach. Remoteness, in turn, was interpreted to mean that contracting parties could only recover that which *both parties could have contemplated at the time of contract.* Sedgwick, supra note 35, at 65–67.

55. Sedgwick, supra note 35, at 69. The remoteness principle, denying recovery for lost profits, was affirmed often by American courts. In Blanchard v. Ely, 21 Wend. 342 (N.Y. Sup. Ct. 1839), the New York Supreme Court of Judicature, the state's highest court at the time, reviewed an action brought for the price of a steamboat. The buyer showed that the machinery of the vessel was defective and that those and other defects caused considerable delay in putting the boat into operation in his established ferry business. Id. at 343. The buyer sought to deduct from the contract price not only the cost of repairs but also the lost profits for the trips that the vessel might have run during the period of the delay, having proved that the net profits lost would be $100. Id. The court disallowed the recovery of lost profits. Id. at 344. In another New York case, a seller contracted to deliver a steam engine to a buyer who owned a mill for manufacturing oil. Freeman v. Clute, 3 Barb. 424 (N.Y. Gen. Term 1848). The seller failed to deliver the machine at the time for performance and the buyer sought the profits it could have made from manufacturing oil had the machine been delivered on time. Id. at 425–26. The court disallowed the claim, limiting damages to the loss of full use of the mill and other machinery and interest on the additional stock purchased in anticipation of the installation of the engine. Id. at 427.

56. Sedgwick, supra note 35, at 57.

57. The Lively, 15 F. Cas. 631, 634–35 (C.C.D. Mass. 1812) (No. 8403).

58. Sedgwick, supra note 35, at 72.

59. Id. at 112.

60. Id.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

gests that *Hadley v. Baxendale*,[61] decided in 1854, may have announced a rule that had already developed in the United States.[62] If so, the primary effect of *Hadley* was not to limit recovery of consequential damages that previous courts would have allowed, as is generally believed, but rather to extend damages recovery by granting the plaintiff the right to recover consequential damages in those instances where he had "communicated special circumstances" to the defendant.[63]

4. *The Merger of the Compensation Principle with the Penalty Rule.* — During the early decades of the nineteenth century, common law courts found themselves with a vexing problem of interpretation. On the one hand, they had to police penal bonds. The restrictions on enforcing these bonds, which initially had developed in courts of equity, were subsequently enacted into statutes in many states.[64] As a result, the general rule evolved that no sum could be recovered on a bond other than that which compensated the plaintiff for his actual loss.[65] On the other hand, liquidated damages clauses were increasingly common terms in executory contracts between commercial parties.[66] Breaching promisors argued that these damages agreements were, in fact, disguised "penalties" and should be subject to the same legal restrictions that had come to narrow the scope of the bonds.[67]

The courts initially seized on the "intent of the parties" as the key to distinguishing liquidated damages from penalties. An invalid penalty was a sum intended only as security for the performance of the executory

---

61. 156 Eng. Rep. 145 (Ex. 1854).

62. In addition to formulating a damages default, the courts during this period developed more clearly the principle that the "contract itself furnishes the measure of damages." Sedgwick, supra note 35, at 200. Mid-nineteenth century contract law thus distinguished and rejected a line of earlier cases that gave the jury wide latitude and discretionary authority to determine the measure of damages, either by reducing or enlarging the award. The amount of compensation was now regulated by the direction of the courts, and the sole object was to ascertain the agreement of the parties, which controlled the measure of damages. Id. Among other things, this principle was an explicit rejection of the concept of breach as "fault." The motives behind the breach were irrelevant. 2 Parsons, supra note 42, at 443.

63. See Scott & Kraus, supra note 19, at 120–27 (discussing the conventional understanding of *Hadley*); Richard Danzig, Hadley v. Baxendale: A Study in the Industrialization of the Law, 4 J. Legal Stud. 249, 279–83 (1975) (same).

64. These statutes had two principal provisions: 1) They gave defendants on bonds for the payment of money the right to pay into court the principal sum of the debt with interest and costs even though the condition was in default; and 2) they required plaintiffs suing to enforce bonds conditioned on a particular performance to assign specific breaches by the defendant and to prove the amount of loss for each breach. Sedgwick, supra note 35, at 393–96.

65. Technically, since the bonds were sealed instruments and thus enforceable on their face, the plaintiff was awarded judgment for the face amount of the bond (since the "penalty" was the debt), but the statute only allowed execution to issue for the amount of the loss actually suffered and proved by the defendant. Id. at 396.

66. See 2 Parsons, supra note 42, at 434–40 (collecting cases).

67. Sedgwick, supra note 35, at 398.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

promise and not intended to be paid. Enforceable liquidated damages, on the other hand, were intended to be paid by the promisor if she elected not to perform the agreement (a termination option, as it were).[68] Complicating matters further was the fact that courts recognized and enforced alternative contracts (well known to the civil law at this time).[69] Promises to do or to refrain from doing an act or, in the alternative, to pay a given sum of money were not considered liquidated damages and, therefore, were not subject to the penalty rule. The understandable result of these diverse holdings was universally decried confusion.

A much applauded "solution" to the confusion of the earlier case law was proposed in *Jaquith v. Hudson* in 1858.[70] In *Jaquith*, the court upheld a liquidated damages clause enforcing a covenant not to compete in a contract for the sale of a partnership interest in a mercantile business. The court explicitly rejected the intent test of enforceability. Instead, the court held that the governing principle reconciling the cases was that damages must be based on "the *principle of just compensation* for the *loss* or *injury actually sustained*; considering it no greater violation of this principle to confine the injured party to the recovery of *less*, than to enable him, by the aid of the court to extort *more*."[71] The compensation principle, the court declared, was a mandatory rule, a "principle of natural justice" and not a default rule, and thus the intention of the parties was irrelevant. Since the compensation principle was the law of the contract, parties were not permitted by express stipulation, however clear the intent, to set it aside.[72]

---

68. See id. at 398–420 (collecting cases).

69. Id. at 398. By mid-century, the following generalizations could be advanced: First, the primary objective of courts in scrutinizing stipulated damages clauses was to determine the true intent of the parties, although the language they used was not conclusive evidence of that intent. Second, in England, there was a presumption in favor of enforcing liquidated damages clauses in executory contracts, while American courts were more reluctant than their English counterparts to admit the agreement of the parties as conclusive. Third, when the agreement was in the alternative, courts would uniformly enforce the obligation to pay money. Fourth, where a court determined that the intention of the parties was to evade the statutory restrictions on penal bonds, the clause would be treated as a penalty and damages limited to actual losses. Fifth, independent of the above, if damages were uncertain and incapable of being measured except by conjecture, then any stipulated damages would be enforceable. Id. at 421–22.

70. 5 Mich. 123 (1858). The "intent of the parties" test that the courts had developed to distinguish between valid liquidated damages and invalid penalties simply did not square with the outcome of the cases. Many courts found an "implied intention" to create a penalty even though the parties insisted otherwise by their express language. Id. at 134–35; see, e.g., Kemble v. Farren, 130 Eng. Rep. 1234, 1237 (C.P. 1829). *Jaquith* purported to reconcile these contradictions.

71. *Jaquith*, 5 Mich. at 133.

72. The court cited no authority for the proposition that the compensation principle of expectation damages was an immutable rule. This is particularly surprising given that this "principle" had only been generally recognized for less than thirty years. The court reverted to maxims rather than to authority, holding that courts could set aside the parties'

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

But if the compensation principle is a mandatory rule, what role, if any, do the parties have in setting damages defaults? The court in *Jaquith* had a ready answer to this question. The task for the parties was to specify just compensation ex ante in those instances where they had a comparative advantage over a court seeking to do so ex post. Such comparative advantage would exist where the provable loss from the breach of the contract was uncertain, remote, or speculative. Thus, while *Jaquith* purported to restrict party sovereignty over stipulated damages, the *Jaquith* rule actually gave greater latitude to nineteenth-century contracting parties than would be true under modern damages rules. Given the then-prevailing view that lost profits could not be recovered because they were too remote or speculative, the anticipated losses in most commercial contracts would be difficult to prove. Consequently, while parties had considerable freedom to stipulate damages under the *Jaquith* rule, it was only to improve the accuracy of compensation and not for other purposes.

5. *From* Jaquith *to the Present: The Expansion of Expectation Damages.* — The 150 years following *Jaquith* reveal a consistent pattern of expansion in the recovery available under expectation damages for breach of contract. In *Globe v. Landa*, Justice Holmes sought to generalize the common law rule of consequential damages as requiring a tacit agreement that the breaching party would be liable for losses caused by special circumstances.[73] But this limitation on consequential damages soon came under substantial pressure. One source of pressure was the increasing sophistication of expert testimony in establishing the quantum of lost profits that might result from breach.[74] Meanwhile, as commercial enterprise expanded, courts came to regard lost profit claims as a natural and direct result of a breach rather than as a special circumstance that required communication.[75] Eventually, the constraint on the recovery of lost profits was limited to the so-called "new business rule," which denied recovery of profits only in the case of a non-established business enterprise.[76]

---

intention on the familiar ground: "*[C]onventus privatorum non potest publico juri derogare* [A private bargain cannot derogate a public right]." Id.

73. 190 U.S. 540, 544 (1903) (explaining that measure of damages "depends on what liability the defendant fairly may be supposed to have assumed consciously, or to have warranted the plaintiff reasonably to suppose that it assumed, when the contract was made").

74. Since the denial of lost profits was based on claims of uncertainty, the ability to establish losses with reasonable certainty justified recovery under an expectation default.

75. See, e.g., U.C.C. § 2-715 cmt. 6 (2003) (explaining that resale is a requirement of which the seller has reason to know).

76. The new business rule is now treated by many courts merely as a presumption in favor of the defendant rather than as an absolute bar to the recovery of lost profits for start-up businesses. See, e.g., Drews Co. v. Ledwith-Wolfe Assocs., 371 S.E.2d 532, 534 (S.C. 1988) ("Modern cases, however, reflect the willingness of this Court and our Court of Appeals to view the new business rule as a rule of evidentiary sufficiency rather than an automatic bar to recovery of lost profits by a new business.").

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

The universal adoption of Article 2 of the UCC has completed the expansion of the expectation damages default. In addition to explicitly rejecting the tacit agreement test and relaxing the restrictions on recovery of consequential damages,[77] the Code (unlike its predecessor, the Uniform Sales Act) specifically invites lost profit claims by volume sellers.[78] Even more significantly, the Code elevates the compensation principle to an overarching norm and a fundamental principle of interpretation.[79] Consequently, most courts under the Code have concluded that, even in thick-market contexts, market damages should not be used where they depart from the economic gain the promisee would have enjoyed had the contract been performed.[80] And, as this default notion of compensation has expanded, the penalty rule has become a tighter constraint on the ability of parties to stipulate damages.[81]

The dominance of the compensation principle is now unquestioned, but the preceding inquiry into its historical roots suggests that the elevation of compensation to a universally applicable norm results more from mistaken path dependence than from a sustained and systematic appreciation of the merits of the rules governing contract damages. In the fol-

---

77. U.C.C. § 2-715(2) & cmt. 2; see also Restatement (Second) of Contracts § 351 cmt. a (1981) (noting the party in breach is liable for loss that "was foreseeable as a probable, as distinguished from a necessary, result of his breach . . . [and] the party in breach need not have made a 'tacit agreement' to be liable for the loss"). But see Unif. Sales Act § 67 (1906) (retaining traditional rule).

78. The Code abandons the common law preference for market damages, which had been enshrined in the Uniform Sales Act, and instead explicitly authorizes the recovery of lost profits against a breaching buyer so as to "eliminate the unfair and economically wasteful results arising under the older law when fixed price articles were involved." U.C.C. § 2-708(2) & cmt. 2.

79. See id. § 1-305(a) (formerly § 1-106) ("The remedies provided by [the Uniform Commercial Code] must be liberally administered to the end that the aggrieved party may be put in as good a position as if the other party had fully performed . . . ."). Section 1-305 seemingly trumps the arguments for measuring recovery other than by ex post expectancy. Scott, Market Damages, supra note 1, at 1169.

80. See cases cited infra note 168.

81. As noted above, the alternative contract doctrine has also been harnessed to the compensation principle in recent years. See discussion supra Part I.A; see also cases cited supra note 23. This development is reflected in the differing treatment of alternative contracts in the first and second Restatements. The first Restatement affirmed the enforceability of alternative contracts and provided for damages in the case of a breach without an election in accordance with the alternative that resulted in the smallest recovery (presumably without regard to whether that recovery was supercompensatory). Restatement of Contracts § 344 (1932). The second Restatement eliminates any specific default rule for damages for breach of an alternative contract. Instead, a comment to the section on liquidated damages and penalties states:

> Although the parties may in good faith contract for alternative performances and fix discounts or valuations, a court will look to the substance of the agreement to determine whether this is the case or whether the parties have attempted to disguise a provision for a penalty that is unenforceable under this Section.

Restatement (Second) of Contracts § 356 cmt. c.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

lowing section, we ask whether the compensation principle is nevertheless theoretically justifiable on economic grounds.

C. *Economic Contract Theory and Compensation*

Given that compensation is now entrenched as the motivating principle behind default damages rules and the regulation of liquidated damages, it is not surprising that scholars have advanced economic justifications for the rules in each case. Nevertheless, the literature in the economic analysis of contracts reflects a growing academic consensus against the penalty rule and raises substantial doubts about the desirability of expectation damages. Indeed, the economics scholarship of the last decade scarcely mentions compensation either as an end or as a means to maximizing the joint welfare of contracting parties.

1. *Expectation Damages and Efficient Breach.* — The principal economic justification for expectation damages is that it compels the promisor to internalize the costs that her breach inflicts on the promisee. The promisor consequently has the incentive to make the efficient breach decision.[82] By internalizing the promisee's loss, the promisor also has the incentive to take the efficient precautions against contingencies that threaten to increase the cost of performance.[83] In addition, expectation damages provide insurance to the promisee against the risk of increases in the promisor's cost of performance that might lead her to breach. This insurance may be efficient if the promisor is risk neutral and the promisee is risk averse.[84]

---

82. See, e.g., Goetz & Scott, Liquidated Damages, supra note 20, at 558–59 (explaining economic justification for rules that allow for efficient breach). As the authors explain:

> Generally, breach will occur where the breaching party anticipates that paying compensation and allocating his resources to alternative uses will make him "better off" than performing his obligation. As long as the compensation adequately mirrors the value of performance, this damages rule is "efficient." It induces a result superior to performance, since one party receives the same benefits as performance while the other is able to do even better. Under the current damages rule, all of these net gains from breaching are retained by the breacher . . . . In order to maintain the efficiency value of the rule, however, it is only necessary that some minimal amount of benefits are retained by the breacher in order to induce him not to perform.

Id.

83. Id. at 579–83.

84. E.g., A. Mitchell Polinsky, Risk Sharing Through Breach of Contract Remedies, 12 J. Legal Stud. 427, 432 (1983); Samuel A. Rea, Jr., Efficiency Implications of Penalties and Liquidated Damages, 13 J. Legal Stud. 147, 152 (1984). To many, the insurance objective seems inapt in contracts for goods or services. Several decades before Justice Holmes wrote his statement recognizing that a contract grants the promisor the option to perform or to breach and pay damages, see supra note 1, at 462, Pollock wrote:

> A man who bespeaks a coat of his tailor will scarcely be persuaded that he is only betting with the tailor that such a coat will not be made and delivered to him within a certain time. What he wants and means to have is the coat, not an insurance against not having the coat.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

Notwithstanding the possible benefits of the current rule, more recent literature has identified a large number of countervailing considerations against expectation damages. Our purpose is not to provide a comprehensive review of these factors, but simply to illustrate that they present a very significant challenge to the case for expectation damages and indeed compensatory damages in general. Expectation damages might not ensure efficient breach if the court cannot accurately and predictably measure the promisee's losses at no cost. For example, even if the courts are accurate on average, a promisor might inefficiently breach a contract with a promisee who values performance more than average, and might inefficiently perform a contract with a below-average promisee. Even if the court can accurately assess the realized damages, the expected costs of doing so may outweigh the incentive and insurance gains from expectation damages, so that the parties may be better off with undercompensatory damages that reduce the likelihood of litigation.[85] At least in part, the common law doctrines of foreseeability and uncertainty may be explained as striking just such a compromise between the goals of minimizing enforcement costs and inducing efficient breach.

2. *The Excessive Reliance Problem.* — Economic contract theory has focused on a more significant problem with expectation damages: They give rise to well-known moral hazard incentives for the promisee to overrely on promises and to fail to take steps to avoid losses from breach.[86] The problem of excessive reliance and insufficient precautions by the promisee exists whenever the damages are compensatory in the sense that they vary at the margin with the losses of the promisee. In addition, when the promisee is insured against breach by an expectation damages award, she lacks the incentive to incur costs that help or induce the prom-

---

Frederick Pollock, Principles of Contract xix (London, Stevens & Sons rev. 3d ed. 1881).

85. If damages are undercompensatory, the stakes in litigation are lower and the parties will therefore spend fewer resources in litigation. Within some range, the resulting saving in litigation cost may outweigh the incremental loss in the incentive and insurance gains caused by agreeing to lower-than-compensatory damages. This tradeoff is discussed more generally in A. Mitchell Polinsky & Daniel L. Rubinfeld, The Welfare Implications of Costly Litigation for the Level of Liability, 17 J. Legal Stud. 151, 152–53 (1988).

86. This is the well-known overinvestment problem that is created by reliance or expectation damages (as well as specific performance). Suppose, for example, that the seller contracts to produce a specialized good for the buyer and the buyer makes a specific investment (one that has no value other than with the good) in order to increase its valuation of the good. The buyer's expectation damages if the seller breaches will be the difference between its realized valuation and the price. The buyer will always receive the return on his investment, even if trade does not occur and the investment is effectively wasted. The buyer's marginal return from investment exceeds the social return, leading to inefficiently high investment. See Steven Shavell, Damage Measures for Breach of Contract, 11 Bell J. Econ. 466, 470–72 (1980) (noting that "in deciding on his level of reliance, [the victim of breach] does not recognize that reliance is in fact like an investment which does not pay off in the event of breach"); see also William P. Rogerson, Efficient Reliance and Damage Measures for Breach of Contract, 15 Rand J. Econ. 39, 47 (1984) (concluding that under expectation damages buyers will choose a greater than efficient level of reliance).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

isor to avoid breach.[87] Several common law doctrines cut back on compensation in order to address this problem. In particular, the mitigation doctrine deprives the promisee of compensation for postbreach losses that she reasonably could have avoided.[88] The foreseeability and uncertainty doctrines remove unusual losses from the measure of damages and thereby dampen the promisee's incentive to rely.[89] Moreover, plaintiffs tend not to be compensated for waiting for payment until after trial, and they do not recover their legal costs, which restores some of the promisees' incentives to protect themselves from breach. But despite these partial correctives, excessive reliance remains a costly feature of expectation damages. To correct for the overinvestment problem, a contract might fix liquidated damages at the expected compensatory level, so that damages do not vary with the promisee's actual expectancy loss.[90] This strategy may result in optimal precautions, but it will not produce efficient "trade or breach" decisions ex post because the realized damages will be either over- or undercompensatory at the time set for performance.[91]

3. *Shift in Theoretical Focus from Efficient Breach to Efficient Investment.* — These analyses of contract remedies suggest that any potential benefits of expectation damages in promoting efficient ex post decisions to trade or

87. See Yeon-Koo Che & Donald B. Hausch, Cooperative Investments and the Value of Contracting, 89 Am. Econ. Rev. 125, 127–28 (1999) (suggesting that forward contracts might undermine the efficiency of cooperative investments).

88. The doctrine of avoidable consequences has both affirmative and negative aspects. The affirmative branch of the doctrine permits recovery of all reasonable expenses the plaintiff incurs in seeking to avoid damages. See, e.g., Rench v. Hayes Equip. Mfg. Co., 8 P.2d 346, 348 (Kan. 1932) (holding that after wrongful termination former employee has duty to mitigate by seeking other employment but may recover costs incurred). The negative branch of the doctrine precludes a plaintiff from passively incurring losses that she could reasonably avoid or from actively increasing such losses where prudence would require an adjustment. See, e.g., Rockingham County v. Lutten Bridge Co., 35 F.2d 301, 307 (4th Cir. 1929) (holding that bridge builder could not obtain damages for work performed after defendant had given notice of cancellation); Restatement (Second) of Contracts § 350 (1981) ("[D]amages are not recoverable for loss that the injured party could have avoided without undue risk, burden or humiliation."). For discussion, see Charles J. Goetz & Robert E. Scott, The Mitigation Principle: Toward a General Theory of Contractual Obligation, 69 Va. L. Rev 967, 973–76 (1983).

89. In general, speculative losses cannot be recouped. These include lost good will, reputation, or emotional distress. Scott & Kraus, supra note 19, at 1036–38. The foreseeability limitation has been considerably eroded over time. Under UCC section 2-715, for example, the plaintiff can recover "any loss resulting from general or particular requirements and needs of which the seller at the time of contracting had reason to know." U.C.C. § 2-715(2)(a) (2003); see also Restatement (Second) of Contracts § 351 & cmt. a ("It is enough . . . that the loss was foreseeable as a probable, as distinguished from a necessary, result of his breach.").

90. Robert Cooter, Unity in Tort, Contract, and Property: The Model of Precaution, 73 Cal. L. Rev. 1, 14–15 (1985) (suggesting that liquidated damages might address both optimal promisor and promisee precautions by imposing "double responsibility at the margin").

91. If the distribution of damages is continuous, any stipulated sum that is set ex ante to reflect expected damages ex post will always turn out to be wrong at the time of breach.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

breach may be offset by the failure of the damages rule to motivate efficient ex ante investment and reliance.[92] One problem is the aforementioned tendency of the promisee to overrely. In recent years, therefore, contract theorists have focused attention on designing contracts that optimize *both* the ex post decision whether to trade or breach and the ex ante incentives to make specific investments.[93] The challenge comes from the fact that the conditions for efficient trade and efficient investment cannot be specified fully in the contract, in light of the transaction costs of foreseeing and describing all possible future states of the world and of verifying the realized state to a court.

Yet, even when information is not verifiable, contract theorists have shown that ex post efficiency can be achieved without expectation damages through a variety of "implementation mechanisms," as long as the parties themselves are symmetrically informed.[94] These mechanisms elicit from the parties their information about the state of the world that can then be used by courts to determine efficient payoffs for that given state.[95] The availability of these solutions diminishes the necessity for

---

92. Aaron Edlin and Stefan Reichelstein resuscitate expectation damages in their model to show that they can lead to efficient investment where only the promisee's investment is relevant and where the parties can renegotiate their agreement. Aaron S. Edlin & Stefan Reichelstein, Holdups, Standard Breach Remedies, and Optimal Investment, 86 Am. Econ. Rev. 478, 478–79 (1996).

93. See, e.g., sources cited supra notes 86–92.

94. Mechanism contracts induce parties to reveal the ex post state truthfully to a court or other decisionmaker. For example, the contract may provide that the buyer must write down how much she values the seller's performance and the seller must write her cost of performance. If the buyer's written valuation exceeds the seller's written cost, then the court will order the parties to trade; otherwise, the court will prohibit trade. It is a dominant strategy under such a mechanism for each party to report the ex post state truthfully. The inability of the parties to renegotiate the consequences of a mismatch is crucial to the revelation mechanism. Some authors have noted that option contracts are message contingent in this way. See, e.g., Ilya Segal, Complexity and Renegotiation: A Foundation for Incomplete Contracts, 66 Rev. Econ. Stud. 57, 57 n.1 (1999) ("[U]nilateral options [ ] make the outcome indirectly contingent on the parties' non-verifiable valuations for the trade by giving them choices ('messages') within the contractual framework."). For discussion, see Robert E. Scott, Rethinking the Default Rule Project, 6 Va. J. 84, 90–93 (2003).

95. See Benjamin Hermalin & Michael Katz, Judicial Modification of Contracts Between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach, 9 J.L. Econ. & Org. 230, 237–41 (1993) (showing fill in the price mechanism as solution to information shortcomings in contracting); Jean Tirole, Incomplete Contracts: Where Do We Stand?, 67 Econometrica 741, 755 (1999) ("[T]he nonverifiability of information by a court is in general no obstacle to the implementation of contracts contingent on this information as long as this information is commonly observed by several parties."). Under some conditions, an option contract might reveal even asymmetrically held private information. Georg Nöldeke & Klaus M. Schmidt, Option Contracts and Renegotiation: A Solution to the Hold-Up Problem, 26 Rand J. Econ. 163, 167 (1995) (showing that option contracts could be conditional on messages between contracting parties).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

compensatory damages to force promisors to internalize the losses caused by breach.

The contribution of expectation damages in promoting efficient breach or decisions to trade is even less valuable if contracting parties are able to renegotiate their agreement. The premise that parties can often renegotiate to efficient ex post outcomes is sufficiently well accepted that contract theorists have largely set aside the concern with efficient breach to focus on the hold-up problems caused by renegotiation.[96] In particular, the division of the renegotiation surplus will not reflect the fact that one or both of the parties have made specific investments contributing to the surplus. Because the specific investments are sunk costs that cannot be redeployed in other projects, the non-investing party can compel a renegotiation of the contract price. Thus the investing party will be compelled to share the payoffs from her investment.[97] In turn, the anticipation of a renegotiation of contractual commitments undermines attempts to establish efficient investment incentives.[98] Given that states of the world and performance obligations cannot be fully described in the initial contract, the courts often cannot discern when a party forcing renegotiation is acting strategically by refusing to carry out an efficient trade. One set of solutions to this problem proposes that the parties create legal or practical obstacles to renegotiation,[99] such as bargaining through or-

96. See, e.g., Oliver Hart & John Moore, Incomplete Contracts and Renegotiation, 56 Econometrica 755, 756 (1988) [hereinafter Hart & Moore, Incomplete Contracts and Renegotiation] (demonstrating hold-up problem in incomplete contracts).

97. For discussion, see Alan Schwartz & Robert E. Scott, Contract Theory and the Limits of Contract Law, 113 Yale L.J. 541, 559–62 (2003).

98. The key to the hold-up threat is the presence of specific investment, whose value in the exchange is shared between the two parties in the course of the renegotiation. See Benjamin Klein et al., Vertical Integration, Appropriate Rents, and the Competitive Contracting Process, 21 J.L. & Econ. 297, 298 (1978) (emphasizing possibility that contracting parties will engage in opportunistic behavior after specific investment is made and appropriable rents are created); Oliver E. Williamson, Transaction Cost Economics: The Governance of Contractual Relations, 22 J.L. & Econ. 233, 234 (1979) ("[O]pportunism is especially important for activity that involves transaction-specific investments in human and physical capital . . . .").

99. See, e.g., Alan Schwartz & Joel Watson, The Law and Economics of Costly Contracting, 20 J.L. Econ. & Org. 2, 17–18 (2004) (discussing circumstances under which very high renegotiation costs protect efficient complex contracts); Tirole, supra note 95, at 760 (same). A precommitment not to renegotiate the contract generally cannot be made simply by contract. Under current law, an agreement not to modify a contract is not enforceable. Restatement (Second) of Contracts § 311 cmt. a (1981); see also Zumwinkel v. Legget, 345 S.W.2d 89, 94 (Mo. 1961) ("The parties to a contract ordinarily are as free to change it after making it as they were to make it in the first instance . . . ."); Beatty v. Guggenheim Exploration Co., 122 N.E. 378, 381 (N.Y. 1919) (Cardozo, J.) ("Those who make a contract can unmake it. The clause which forbids a change may be changed like any other . . . . Whenever two men contract, no limitation self-imposed can destroy their power to contract again."); Christine Jolls, Contracts as Bilateral Commitments: A New Perspective on Contract Modification, 26 J. Legal Stud. 203, 208–09 (1997) (explaining that provisions in a contract constraining modification of the contract are not enforceable).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

ganizations whose internal regulations require many layers of consent. A more popular theoretical approach is to try to predetermine the allocation of the renegotiation surplus by effectively assigning bargaining power to one party or the other.[100]

In sum, contemporary economic contract analysis indicates that expectation damages do not stimulate efficient ex ante investment and are neither sufficient nor necessary to achieve efficient ex post trade.[101] Parties have the choice between two types of contracts: 1) more complete contracts that yield both efficient trade and efficient investment but can only succeed if there is no prospect of renegotiation; or 2) simple contracts that the parties may renegotiate to achieve efficient trade without compromising investment efficiency.[102] This theoretical literature also raises doubts about the doctrinal constraints on liquidated damages by showing that penalties may serve to improve the efficiency of specific investment by ameliorating the hold-up problems discussed above.[103]

---

100. E.g., Philippe Aghion et al., Renegotiation Design with Unverifiable Information, 62 Econometrica 257, 258 (1994); Tai-Yeong Chung, Incomplete Contracts, Specific Investments, and Risk Sharing, 58 Rev. Econ. Stud. 1031, 1035 (1991); Nöldeke & Schmidt, supra note 95, at 165. The idea in these three articles is to set the correct investment incentives for one party through the setting of the default point in renegotiation and for the other party by assigning her bargaining power over the division of the renegotiation surplus. These models rely on the assumption that the parties can specify bargaining power in a way that is immune to renegotiation. Property ownership rights may also be used to determine the parties' bargaining positions in ex post renegotiation. Sanford J. Grossman & Oliver D. Hart, The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration, 94 J. Pol. Econ. 691, 716–17 (1986); Oliver Hart & John Moore, Property Rights and the Nature of the Firm, 98 J. Pol. Econ. 1119, 1152 (1990). Yet another approach is to permit bargaining power in renegotiation to be exogenously determined, but to set the contract quantity so as to balance the likelihood of renegotiation to a lower or higher value, in such a way as to also cancel out under and overinvestment tendencies. Edlin & Reichelstein, supra note 92, at 479.

101. But see Eric Posner, Economic Analysis of Contract Law After Three Decades: Success or Failure?, 112 Yale L.J. 829 (2003). As Posner observed:

> [It is possible] that on average pre-performance investment is not a significant issue, or, if it is, it is adequately controlled by the doctrine of mitigation. The rule of expectation damages is optimal because the perform-or-breach decision matters most, with specific performance reserved for cases where valuation problems are insurmountable. But this view is unsupported by any evidence.

Id. at 839.

102. See, e.g., Schwartz & Watson, supra note 99, at 17–19; Tirole, supra note 95, at 743–44.

103. In the next part, we attack the penalty rule from a different perspective by demonstrating that it impedes the sale of insurance from promisees to promisors. See infra Part II.C.2. An excellent review of the arguments for and against the penalty rule may be found in Aaron S. Edlin & Alan Schwartz, Optimal Penalties in Contracts, 78 Chi.-Kent L. Rev. 33 (2003). Proponents of the penalty rule suggest it prevents contracting parties from agreeing to inefficiently high damages in order to extract a larger portion of the surplus of a new entrant who arrives during the contract term. See, e.g., Philippe Aghion & Patrick Bolton, Contracts as a Barrier to Entry, 77 Am. Econ. Rev. 388, 389 (1987); Tai-Yeong Chung, On the Social Optimality of Liquidated Damage Clauses: An Economic Analysis, 8 J.L. Econ. & Org. 280, 281 (1992). But the possibility that the parties

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

D. *Commercial and Consumer Practices of Termination*

Contracting parties invoke a range of techniques to contract away from the compensation principle, with varying degrees of success. They may frame remedial provisions as substantive terms such as a cancellation right.[104] The parties may contract for alternative methods of contract performance rather than providing for a primary obligation to perform and a secondary obligation to pay damages.[105] Alternatively, one party may grant the other an explicit option[106] or the right to cancel upon

---

may use penalties to deter the entry of competitors into the seller's market does not, by itself, justify the current penalty rule. Under current law, courts strike what they perceive to be penalty terms whether or not those terms were used to impede entry. A party should always be free to argue that any term would create a negative externality or perpetuate a market failure. But it is a mistake to treat as a sufficient proxy for these inefficiencies a liquidated damages clause that would overcompensate the promisee in expectation.

104. See, e.g., Restatement (Second) of Contracts § 368(1) ("Specific performance or an injunction will not be granted against a party who can substantially nullify the effect of the order by exercising a power of termination or avoidance."). In contrast, section 361 provides that the court can order an injunctive remedy in the face of a liquidated damages term.

105. Traditional analysis has distinguished between the following terms: 1) the alternative provision designed to secure performance of the primary promise (a liquidated damages clause); and 2) two promised alternatives between which the promisor can choose, each an agreed exchange for the consideration given by the promisee. Restatement of Contracts § 339 cmt. f (1932); see also, e.g., Garrett v. Coast & S. Fed. Sav. & Loan Ass'n, 511 P.2d 1197, 1201 (Cal. 1973) (recognizing "validity of provisions varying the acceptable performance under a contract upon the happening of a contingency" but reviewing clause as liquidated damages provision when it "is in fact a contract contemplating but a single, definite performance with an additional charge contingent on the breach of that performance"). The former is at risk of being found unenforceable as a penalty, while the latter would be an enforceable alternative contract. The first Restatement, in a classic understatement, acknowledges that enforceable alternative contracts may be confused with invalid liquidated damages provisions. Restatement of Contracts § 344 cmt. c. Modern courts have been inclined to bring compensation to bear in the decision to enforce even alternative promises. See cases cited supra note 23.

106. For example, a unilateral promise invites the promisee to accept by rendering a performance rather than by a promissory acceptance. The promise creates an option that is exercised once the promisee begins the invited performance. Restatement of Contracts § 45(1). See generally Katz, supra note 1 (manuscript at 14) ("Structuring a contract as an option may in many cases help the parties evade the penalty doctrine . . . ."). If the parties fail explicitly to characterize the contract as an option, the agreement is vulnerable to challenge as a penalty even though it functions as an option. See, e.g., Wasserman's Inc. v. Township of Middletown, 645 A.2d 100, 108–11 (N.J. 1994) (concluding that township's right to cancel lease upon payment of percentage of lessee's gross receipts is subject to reasonableness review as stipulated damages provision). For discussion of *Wasserman* and the option characterization, see Victor P. Goldberg, Framing Contract Law: An Economic Perspective (forthcoming 2005) (manuscript at pt. 5.3, on file with the *Columbia Law Review*) [hereinafter Goldberg, Framing Contract].

Options that appear to be "free" are sometimes challenged on the grounds that the purported contract lacks mutuality and thus is unenforceable. In response, courts often enforce the option on the condition that its exercise is subject to a good faith or reasonableness limitation. See, e.g., Mattei v. Hopper, 330 P.2d 625, 627 (Cal. 1958) (holding that good faith limits discretion to exercise option to cancel real estate contract if

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

payment of a fee or loss of a deposit. A requirements or output contract permits one side to avoid the obligation to purchase or deliver larger quantities of goods.[107] In view of these substantive alternatives, breach damages are best seen as a subset of the larger category of termination provisions. These terms share the common feature that one party has the option to terminate the contemplated exchange by paying some amount or incurring some cost. That cost is unrelated to compensation. In practice, this termination fee is often undercompensatory (consider, for example, the number of buyers who can return goods for free with no questions asked). But sometimes buyers agree to make payments that are likely to be overcompensatory (both ex post and ex ante). Consumers, for example, pay cancellation fees for walking away from airline tickets or concert performances even when their seats are resold.

The success of these techniques in avoiding the regulation of liquidated damages is far from certain because of the broad reach and prominence of the compensation principle. We suspect that many contracting parties are discouraged from contracting for efficient termination rights by the dominance of the compensation principle in contract law. As noted in Part I.B, the penalty rule is a formidable obstacle.[108] The

---

lease satisfactory to buyer cannot be obtained). For a critique, see Goldberg, Framing Contract, supra (manuscript at pt. 2).

107. See U.C.C. § 2-306(1) (2003) (barring, in requirements or output contracts, demands for or tenders of any "quantity unreasonably disproportionate to any stated estimate or . . . to any normal or otherwise comparable prior output or requirements"). Discretion under an output or requirements contract is effectively constrained only by the duty of good faith. Id. A number of courts have recognized that underdemand in the case of a requirements buyer (or undersupply for an output seller) need only be justified by a "valid business reason" to pass the test of good faith imposed by UCC section 2-306(1). See, e.g., Empire Gas Corp. v. Am. Bakeries Co., 840 F.2d 1333, 1340–41 (7th Cir. 1988) (suggesting that buyer's burden of producing "valid business reason" for reducing its requirements is minimal). Thus construed, a requirements contract can be characterized as an unconstrained option to take as small a quantity of the good as the buyer chooses. The buyer's discretion is regulated in the overdemand situation to prevent the buyer from opportunistically exploiting the contract price and reselling the goods in a favorable market. Id. at 1338; see also Goldberg, Framing Contract, supra note 106 (manuscript at pt. 2.4) (discussing various doctrinal solutions to the overdemand problem).

Many other contracts grant one of the parties discretion to determine a significant exchange term. For example, the seller under an installment contract has discretion over the content of any installment. Breach does not occur unless nonperformance substantially impairs the value of the whole contract. U.C.C. § 2-612(2) to (3). UCC default rules also grant the buyer an option relating to assortment of goods and the seller an option relating to shipment. Id. § 2-311(2).

108. Two of our students undertook a search of all federal and state appellate cases decided between January 1, 1988 and January 1, 2004 that invoke liquidated damages and the penalty rule. The search returned 333 recorded decisions where enforceability was at issue. A random sample of 109 of these cases was examined in detail. In 37% of the sample cases, the courts found that the stipulated damages provisions in question were void as penalties. Jason K. Binder & Michael Labriola, An Empirical Analysis of Liquidated Damages Clause Jurisprudence in the United States from 1988 to 2003, at 18 (2004) (unpublished seminar paper, University of Virginia School of Law) (on file with the

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

linkage of a quasimandatory compensation principle with both the penalty rule and the expansive default of full expectation damages may have resulted from the anomalies of the penal bond and the early common law forms of action. But once entrenched, these rules have proven to be remarkably durable despite the claims of scholars that they preclude parties from writing efficient contracts (or at least constrain them in doing so). Prepayments of contract price, stipulated damages in licensing contracts, and nonrefundable deposits are vulnerable to being struck down under the penalty rule.[109] Moreover, as we noted earlier, many courts scrutinize even alternative contracts such as take-or-pay provisions to ensure they have some link to the compensation of the seller.[110] Although courts have been less inclined to strike down undercompensatory damages provisions and termination rights than penalties,[111] doctrinal statements ex-

---

Columbia Law Review). It is important to emphasize the significance of the roughly 40% rate with which contemporary courts refuse to enforce liquidated damages clauses. Assuming that none of the litigated contracts were unconscionable (there were no such holdings in the sample), then under accepted contract principles the parties should have been free to stipulate damages as they saw fit. Thus, the appropriate baseline from which to assess these findings is to assume that, but for the penalty rule, 100% of the contested clauses would be upheld as enforceable. Moreover, since lawyers drafting liquidated damages clauses are likely cognizant of the penalty rule, one would expect them to attempt to draft these provisions so as to pass the legal test for enforceablity. Thus, the fact that nearly 40% of the litigated damages clauses fail the test is strong evidence of the continued vitality of the penalty rule.

109. See, e.g., MCA Television Ltd. v. Pub. Interest Corp., 171 F.3d 1265, 1275 (11th Cir. 1999) (invalidating as a penalty a clause awarding nonbreaching party full payment of entire contract price plus right to revoke breacher's license to broadcast); Shree Ganesh, Inc. v. Days Inns Worldwide, Inc., 192 F. Supp. 2d 774, 786 (N.D. Ohio 2002) (invalidating liquidated damages clause in franchise agreement); In re Admetric Biochem, Inc., 284 B.R. 1, 9–11 (Bankr. D. Mass. 2002) (invalidating deposit-plus-acceleration clause in commercial lease where it constituted 20% of total rent reserved under lease).

110. See supra note 23 and accompanying text. For discussion of take-or-pay clauses, see Victor P. Goldberg, Bloomer Girl Revisited or How to Frame an Unmade Picture, 1998 Wis. L. Rev. 1051, 1074; Goldberg, Discretion, supra note 1, at 339–47; Goldberg, Net Profits Puzzle, supra note 1, at 539.

111. See supra note 21 and accompanying text. Thus, for example, commercial sellers have designed various limitations of remedy provisions such as the ubiquitous "repair and replacement" clauses. Richard Epstein observes that the warranty provisions in many sales contracts limit the buyer's remedies for breach either to liquidated damages that are undercompensatory or to the repair or replacement of the damaged goods at the option of the seller. Richard A. Epstein, Beyond Foreseeability: Consequential Damages in the Law of Contract, 18 J. Legal Stud. 105, 114–21 (1989). Epstein also describes the standard limitation on damages in delivery services such as Federal Express, which undercompensate even the average consumer. Id. at 120–21. Consistent with the analysis summarized above, Epstein concludes that

> [a]gainst the backdrop of express contractual provisions, there is ample reason to doubt that the expectation measure of damage of the classical common law maximizes the joint gains of the parties ex ante. If it did, we should expect to observe it frequently in practice, which is decidedly not the case.

Id. at 116.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

plicitly maintain that underliquidation is subject to the same constraints as overliquidation.[112]

II. Embedded Options

The preceding discussion underscores the importance of understanding how contracting parties make their choices among alternative termination provisions and, specifically, how they select a termination fee. In this Part, we characterize termination provisions as options embedded in contracts and the termination fee as effectively the price paid for the option enjoyed by the promisor. We focus on the option created by the termination right held by a buyer of a good, and assume accordingly that the seller's obligation is specifically enforceable. In the case of termination by breach, this option price is the amount of damages owed by the breaching buyer. We then examine the factors that determine the option price to which the parties would agree. Although the protection of specific investment remains an important contract objective, our analysis shifts the focus to risk management. We emphasize that risk management is a widely pursued objective, even by risk-neutral parties, and that embedded options in thin-market contracts (such as sales of goods) are important components of risk management strategy. We conclude that contracts set option prices equal to either the ex ante or ex post loss from termination only where neither risk management nor promisee-specific investment are important contracting objectives, thereby raising serious doubts about the compensation principle in the measure of contract damages.

A. *Termination Rights and Call Options*

Scholars who frame contract breach in option terms typically analyze the buyer as having purchased the good for the contract price, P, and a put option on the good with exercise price, x.[113] In this paper, however, we prefer the equivalent characterization in which the buyer pays an *option price*, d, to acquire a *call option* to purchase the good with an *exercise price*, x. The sum of the option price and the exercise price is the contract price, $P = d + x$. If the buyer's promise is enforced by damages, then

112. See U.C.C. § 2-718 cmt. 1 ("A term fixing unreasonably large liquidated damages is expressly made void as a penalty. An unreasonably small amount would be subject to similar criticism and might be stricken under the section on unconscionable contracts or clauses.").

113. See supra note 2. The payment of the termination fee or damages by the buyer or seller, respectively, is subject to a further option if there is a risk of insolvency. Each party can obtain a discharge from the obligation by surrendering his or her assets in a bankruptcy proceeding. Thus, the option enjoyed by a contracting party is essentially a compound option because it includes both the termination option in the contract and the bankruptcy option held by all debtors. Triantis, Insolvency and Bankruptcy, supra note 1, at 682–87. For the purposes of this Article, however, we assume that both parties cannot avoid their debts by bankruptcy or dissolution.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

d is the damages liability and x is the difference between the contract price and those damages.

We prefer the call option characterization because it is more analytically revealing than the contract-plus-put combination. Consider, for example, a contract under which the buyer purchases a widget for $15 and can recover $11 if she returns it within thirty days.[114] Under the put-call parity rule, the buyer's position can be characterized either as 1) the combination of the widget together with a put option with exercise price $11; or 2) a call option on the widget with exercise price $11.[115] By using the call option characterization, we can describe the option by the pair (d, x) that isolates the price of the option in d: The buyer pays a price of $4 for a call with exercise price $11. If we used the alternative characterization, the price of the put would be embedded in the contract price and thus could not be isolated and compared to contract damages (the contract price of $15 in the example above reflects the value of the put together with the value of the widget).

The parties can choose from a set of alternative option contracts, (d, x). The values of d and x will tend to be inversely related, and the contract price will be higher the easier it is for the buyer to walk away from the contract.[116] Consider first a contract in which the buyer agrees to buy a widget for $12 with no termination right, (12, 0). Now, suppose the seller offers the buyer an alternative contract under which the buyer makes a deposit, d, and holds an option to purchase the widget for $1.

---

114. For now, set aside the question of which party has possession of the widget during the thirty-day period (by assuming that use value and depreciation are trivial) and the question of who holds the $15 during that period (by assuming that the market discount rate is zero). We take up these questions in Part II.D.1.

115. Specifically, a call option is equivalent to the combination of holding the underlying asset and a put option on the same asset and borrowing the exercise price. E.g., Hans R. Stoll, The Relationship Between Put and Call Option Prices, 24 J. Fin. 801, 805–10 (1969). We assume a zero discount rate.

116. Some economic models of contracting analyze a contract that provides for two prices to be paid by the buyer. The buyer pays p_1 if there is trade and p_0 if there is not, where p_0 might be negative. The critical feature is the difference, $p_1 - p_0$, and the parties are otherwise free to set p_0 where they wish. See, e.g., Hart & Moore, Incomplete Contracts and Renegotiation, supra note 96, at 760–61 (explaining that efficient trade will take place if $p_1 - p_0$ is greater than seller's cost for the item and less than buyer's valuation of the item); Nöldeke & Schmidt, supra note 95, at 165 (arguing that by varying $p_1 - p_0$, simple option contracts can give "[promisors] the right incentives to invest"). "[O]nly the difference $p_1 - p_0$ matters for efficiency. Therefore, we have one degree of freedom in choosing p_0 which can be used to share the expected surplus *ex ante* between the two parties." Klaus M. Schmidt, Contract Renegotiation and Option Contracts, *in* 1 New Palgrave Dictionary, supra note 27, at 433. In the options analysis of property-versus-liability rules, Ian Ayres and Paul Goldbart analyze a similar choice among combinations of payments that might be made between plaintiffs and defendants in, for example, nuisance suits. By "convexifying" the available combinations, they enable the court to choose among distributional results without affecting allocative efficiency. Ian Ayres & Paul M. Goldbart, Optimal Delegation and Decoupling in the Design of Liability Rules, 100 Mich. L. Rev. 1, 28–33 (2001).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

The buyer pays d to acquire a call option with exercise price \$1. The deposit, d, is effectively the price the buyer pays for the option. If there is some possibility the buyer will value the widget for less than \$1, she will be prepared to pay a deposit—effectively an option price—greater than \$11. Writing this option may also be costly for the seller,[117] but the parties will choose to trade the option if this cost is less than the value of the option to the buyer. If the seller has some bargaining power, she will be able to capture some of the surplus. It is therefore likely that the aggregate contract price will be greater than \$12 when the seller writes a call option with an exercise price of \$1. As a general matter, the sensitivity of the option price to changes in the exercise price is such that, for each dollar increase in the exercise price, the price of the option decreases by less than a dollar. Therefore, the aggregate contract price rises asymptotically with the exercise price as the parties approach $(0, \hat{P})$, where $\hat{P}$ is the maximum value for the contract price.[118]

If we begin our analysis instead at the $(0, \hat{P})$ end of the spectrum, suppose that $\hat{P} = x = \$18$. The seller effectively has given the buyer a free call option with an exercise price of \$18. We later discuss the possible motivations for giving such a free option.[119] For now, note that if the seller offers to reduce the exercise price from \$18 to \$17, the buyer will be willing to pay an option price in the form of a deposit—but in an amount less than \$1. Similarly, if the seller offers to reduce the exercise price further, the buyer will be prepared to increase her deposit, but by less than the reduction in the exercise price. Of course, the seller does not seek to maximize the contract price per se, but rather to maximize the surplus between the option price and the cost of writing the option.

The ways in which the price of the call option, d, is paid are as diverse as the forms of termination fees. The buyer may make a nonrefundable deposit, agree to a cancellation fee, or assume a commitment to pay damages in the event of breach. The common feature is that the buyer pays for the call option by incurring an initial cost specified by contract and can subsequently choose to incur an additional cost to execute the contract exchange.[120] If the option price is in the form of a

117. If, as we assume, the seller has no termination rights, the seller's cost includes the fact that the seller cannot entertain better offers from other buyers.

118. In nominal (dollar) terms, the option price is most sensitive to changes in exercise price when the option is near or at the money (exercise price equal to the buyer's valuation).

119. These free options are common and might be explained as cases where the seller incurs a trivial cost in writing the option in a competitive market (e.g., a reservation at a restaurant with a heavy street clientele) or where the seller uses the free option as a marketing tool (e.g., in the place of a free gift to potential customers). See also infra note 151 and accompanying text. Victor Goldberg suggests that, in some cases, sellers grant options in order to acquire information from their buyers. Goldberg, Framing Contract, supra note 106 (manuscript at pt. 2.2).

120. Recall that the buyer is assumed to remain solvent at all times. See supra note 113. The seller's option to terminate can be similarly presented. If the seller has a termination right, then she effectively incurs a liability for the termination fee and can pay

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

commitment to pay damages upon breach, then this price depends on the materialized loss from breach suffered by the seller. Consequently, the buyer faces a distribution of option prices during the term of the option. This uncertainty does not change the analysis. The option to breach and pay damages conventionally may be termed an exchange option: the right to exchange one asset of uncertain value (the damages liability) for another (the payoff from a completed exchange). For our expositional purposes, however, we will stick to the call option characterization whether the option price is fixed or uncertain at the time of contract.

Indeed, there may be other sources of uncertainty in the option price. The enforcement of breach damages may be less than perfect or the duration of the option may be uncertain. For example, many consumers purchase retail items where the return policy is unclear; retailers may be deliberately vague. The consumers' ability to walk away from the deal may need to be discounted, therefore, according to the probability that the retailer will permit a refund of varying amounts, and according to the distribution of negotiated or judicial outcomes in the event of any dispute. Practically, these factors are significant in many circumstances. We suggest in Part II.B.1 below that the buyer may be willing to pay a premium to avoid these risks. This premium may be in the form of liquidated damages that will be struck down by a court—inappropriately, in our view—as supercompensatory and a penalty.

In the discussion that follows, we distinguish between two sources of value of embedded call options: 1) preserving the real options given by nature (without the need for renegotiation between the parties); and 2) insuring the buyer against the risk of fluctuations in her valuation of, and in the cost of, the seller's performance. Consistent with conventional efficient breach theory, real options analysis suggests that expectation damages lead to the efficient exercise of real options. As we noted earlier, however, the embellishments to contract theory over the past twenty-five years have emphasized the significance of specific investments and the ability of parties to renegotiate their contracts. As a result, it is now clear that expectation damages are neither necessary nor sufficient for efficiency. The analysis that follows introduces important and widely applicable risk management considerations indicating that optimal termination provisions are highly context contingent and are determined by the factors of insurance and market conditions, rather than the compensation principles of contract law.

---

the additional cost of delivering the good to the buyer in order to receive the contract price. The case of bilateral buyer and seller's options is substantially more complex. See Triantis & Triantis, supra note 1, at 184–94 (explaining that in bilateral options the maturity date of one party's option is determined by the exercise of the other party's option thereby "mak[ing] it more difficult for a party to value the remaining life of any option at any point in time and therefore to decide whether to repudiate").

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

B. *Real Options, Embedded Written Options, and Insurance*

Contracts can embed real options and written options. The distinction between the two is important. The exercise of real options yields value to the holder but imposes costs on no one. In other words, no party *writes* real options; they are the product of nature. For example, a firm may have flexibility in choosing when to exploit a given opportunity. It can delay its investment in the project and hold the option to invest at a later date. Once it begins investing, the firm then holds options to accelerate, decelerate, or even abandon the project. Consider the ability to delay the investment, effectively a call option on the project. This option is valuable because the firm has the opportunity to take action to limit its losses as it receives new information about project payoffs. When a venture is carried out by contract rather than in an integrated firm, the parties may agree to termination rights in order to preserve the firm's valuable real options. The termination rights assign the exercise decision to one or both of the parties, depending on their comparative access to new information. At the same time, the contract should ensure that the holder of the option has the appropriate incentives to make efficient exercise decisions.[121] Consistent with conventional efficient breach theory, expectation damages promote efficient option exercise by compelling the decisionmaker to internalize the costs borne by the other party. Thus, only if the optimal exercise of real options were relevant would the parties be inclined to agree to expectation damages and to limit their respective termination rights accordingly.

A second type of call option is not given by nature, but instead written by one party (in our analysis, the seller) in favor of the other (the buyer). The seller who writes an option bears a cost when the buyer exercises it (the option is said to be "in the money"). Thus, whenever a termination provision is not ex post compensatory, the seller will have written an option in favor of the buyer. Such written options are efficient if their ex ante value to the buyer exceeds the ex ante cost to the seller, and this may be the case because of the comparative risk-bearing advantage of the seller. The presence of these written embedded options explains why many contracts have termination provisions whose prices deviate from the compensation principle.

1. *The Insurance Function of Written Options.* — When insurance concepts are invoked in the analysis of contract damages, they typically concern the risk of the promisor's breach. Thus, a buyer may be willing to pay a premium to be insured against increases in the seller's costs that lead the seller not to perform, and the seller may be willing to accept that

121. The virtue of options in harnessing the private information of the option holder has been developed more generally in evaluating the benefits of liability rules enforcing legal entitlements. See Louis Kaplow & Steven Shavell, Property Rules Versus Liability Rules: An Economic Analysis, 109 Harv. L. Rev. 713, 725 (1996) (arguing in favor of a liability rule because "it allows the state to harness the information that the injurer naturally possesses about his prevention cost").

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

risk in return for the premium.[122] There are several reasons why this trade may be advantageous to both parties. Most obviously, the buyer may be risk averse while the seller is risk neutral. In addition, the seller may enjoy a comparative advantage in reducing the risk of increases in its costs and thus may be better able to manage the risk for the buyer. This much is well known. But the buyer bears two further risks that have thus far been neglected in the economic analysis of contract: 1) the risk of changes in the *value* to the buyer of the seller's performance; and 2) the risk of fluctuations in the seller's profit that may increase the buyer's liability for breach damages. The buyer, therefore, may be willing to pay a further premium in order to be able to shift both the risk in the value of performance and the risk in the seller's costs. Once again, the parties will agree that the seller insure some or all of these risks either because the buyer is more risk averse or because the seller has a comparative advantage in managing the risk of changes in its costs or in the value of its performance to the buyer.

To illustrate this point, consider a venture that entails the production of a good at a cost, $\tilde{c}$, and value, $\tilde{v}$, where both variables are initially random. An integrated firm with this opportunity bears exogenous risks in the joint distribution of $\tilde{v}$ and $\tilde{c}$ that cause $\tilde{v}$-$\tilde{c}$ to rise or fall. The firm might preserve its option to be able to walk away if subsequent new information reveals that $\tilde{v}' < \tilde{c}'$. Now, consider that two parties might pursue the same project by a fixed-price contract with price, P: The seller produces the good at a cost, $\tilde{c}$, and delivers it to the buyer who values it at $\tilde{v}$. Under a specifically enforceable contract, the seller bears the cost risk, $(P - \tilde{c})$, and the buyer keeps the value risk, $(\tilde{v} - P)$. The parties can preserve the real option that would be held by the integrated firm by choosing to enforce their contract by expectation damages. Each party shares in the value of this real option by being able to shed some of her respective downside risk (in particular, the risk that $\tilde{v}$ falls at the same time as $\tilde{c}$ rises) without passing any additional risk to the other party.[123] When the parties agree to expectation damages, the seller does not write an option for the benefit of the buyer; it is given by nature.

Beyond the preservation of real options, however, the seller might also *write* an option in favor of the buyer that would further insure the buyer against the risks of the project. When the seller writes the option, the termination fee is not equal to the seller's realized loss from termination (i.e., expectation damages). Assume that the seller will always specifically perform (i.e., deliver the goods). Under our notation, a regime of

122. E.g., A. Mitchell Polinsky, An Introduction to Law and Economics 63–66 (3d ed., Aspen 2003). Like other commentators, Polinsky assumes that the seller's costs fluctuate (because of the risk that a new buyer will emerge to offer a higher price for the goods), but that the buyer's valuation is constant.

123. This is the real option described in the previous section, whose value depends on the absence of perfect positive correlation between the distributions of $\tilde{v}$ and $\tilde{c}$. Neither the seller nor buyer suffers from conditions that increase $\tilde{c}$ and decrease $\tilde{v}$, so that $v < c$.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

expectation damages gives the buyer a call option, $(d, x) = (P - \tilde{c}, \tilde{c})$; that is, the buyer incurs a liability of $P - \tilde{c}$ and can exercise its call by paying the balance of the contract price, $\tilde{c}$. Note that both the option price and the exercise price are uncertain at the time the parties enter into their contract, and these amounts will be judicially determined at a later date. Consequently, the buyer may be prepared to pay a premium in order to avoid the risk in the option price that is due to fluctuations in $\tilde{c}$. The seller may therefore write a call option with a fixed price (i.e., a liquidated damages provision) that the court would perceive as supercompensatory in the ex ante sense. In other words, rather than $(P - \tilde{c}, \tilde{c})$, the parties may agree to $(E(P - \tilde{c}) + k, \mathrm{E}(\tilde{c})) = (P - \tilde{c}\, E(\tilde{c}) + k, \mathrm{E}(\tilde{c}))$, where $\mathrm{E}(\tilde{c})$ is the expected value at the time of contracting and k is the premium paid by the buyer.[124]

Finally, recall that the parties also have the choice among different combinations of option price and exercise price, (d, x). By raising the exercise price, x, and reducing the option price, d, below the seller's expected profit, $P - E_0(\tilde{c})$, the parties can shift a larger portion of the valuation risk, $\tilde{v}$, from the buyer to the seller. In this way, the buyer can escape more of the lower tail of the distribution by walking away from her call option when her valuation falls below the contract price. Of course, the buyer in return must agree to a higher exercise price. But if a court interprets the option price, d, as stipulated damages imposed in the event of the buyer's breach, these damages will appear to be undercompensatory even though they are a fair price for the call option with that exercise price.[125] The discrepancy arises from the fact that compensation measures the seller's lost payoff given that there is no exchange. The option price takes into account the exercise price that the seller will receive in those states of the world in which the option is exercised. From this perspective, true damages more nearly approximate the seller's cost of writ-

---

124. This may explain why passengers pay significant penalties when they change their minds about flights they purchase from airlines. These penalties are premiums paid by travelers for insurance against declines in their valuation of booked flights (for example, where the purpose of the trip was a wedding that is cancelled), but they can lead to inefficient trade when the traveler values the flight less than a potential replacement passenger. See infra text accompanying note 143.

125. If we think briefly about the seller's option, we note that liquidated damages payable by terminating sellers are often pegged significantly below the compensatory level. See supra note 111. For example, repair and replacement are common remedies for defective performance. Contracts also limit recovery for packages lost by couriers or luggage lost by airlines. Recently, distributors of electricity have issued interruptible electricity contracts that allow the distributor some flexibility to interrupt electric service to commercial customers, which they exercise in times of demand spikes. Ross Baldick et al., Valuation and Optimal Interruption for Interruptible Electricity Contracts 1 (2003) (unpublished manuscript, on file with the *Columbia Law Review*). Some contracts provide for financial compensation at the time of the interruption, but others provide for a payment in advance. Id. at 4–5. Neither payment purports to reflect the loss suffered by the customer.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

ing the option than the seller's loss when the option expires without exercise.

2. *The Advantages of Risk Management.* — The preceding discussion argued that the seller will agree to write an embedded option in favor of the buyer where the seller can better manage contract risks. Although some scholars believe that risk aversion can explain very little of the contracting that occurs in practice,[126] in fact corporate managers actively hedge interest rate, exchange, or commodity price risks; and risk management is the topic of much literature both by serious academics and practitioners.[127]

Managing risk at the firm level has a number of advantages that cannot be replicated by investors and that might therefore lower the firm's cost of capital.[128] Under a progressive tax regime, a firm's tax liability is lower if income is stable rather than volatile. In addition, the risk aversion of managers raises the cost of realigning their incentives by increasing the portion of compensation they derive from stock or stock options (particularly restricted stock). By dampening the variance in firm profitability, risk management reduces the cost of such compensation and allows firms to load up on incentive compensation.[129] Moreover, when exogenous risks are managed, performance-based compensation targets more accurately the skill and effort of managers. Risk management also encourages shareholders to concentrate their investments in the firm rather than diversifying their portfolios, thereby improving their incentives to monitor management. In the same vein, risk management

126. See, e.g., Victor P. Goldberg, Aversion to Risk Aversion in the New Institutional Economics, 146 J. Institutional & Theoretical Econ. 216, 220–21 (1990) (arguing that focus on risk aversion neglects the fact that "[p]eople who really are risk averse can act as if they are risk neutral in some contexts and risk loving in others"). But see id. at 218 (suggesting that insurance questions turn on whether a firm should outsource the task of minimizing likeliness of adverse contingency). Although risk aversion may be accurate in describing consumers, it may seem inappropriate where business entities are concerned, particularly in light of capital market asset pricing theory. This theory predicts that managers of business firms cannot increase firm value by buying insurance because investors are at least as capable of hedging or diversifying nonsystematic firm risks in their financial portfolios. Moreover, the theory also holds that firms cannot create value by hedging systematic risks because shareholders can choose their desired portfolio risk by altering their asset allocation between risky and risk-free assets. See, e.g., Richard A. Brealey & Stewart C. Myers, Principles of Corporate Finance § 7.5 (6th ed. 2002) (stating that diversification adds no value to a firm since investors can more easily diversify than can firms). But see source cited infra note 128 and accompanying text.

127. See, e.g., René Stulz, Mastering Risk I: Diminishing the Threats to Shareholder Wealth, Fin. Times, Apr. 25, 2000, at 8.

128. For a summary of the empirical articles supporting this proposition, see Charles W. Smithson & Clifford W. Smith, Jr., Managing Financial Risk: A Guide to Derivative Products, Financial Engineering and Value Maximization 505–09 (1995).

129. E.g., René Stulz, Optimal Hedging Policies, 19 J. Fin. & Quantitative Analysis 127, 136 (1984) ("[I]f the manager is not allowed to hedge, and if this restriction implies that the instantaneous variance of his or her income is higher than otherwise, he or she will require a higher expected income . . . to accept employment.").

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

reduces the conflict between stockholders and debtholders over capital budgeting and investment decisions. If the probability of insolvency is high, shareholders are more likely to induce their managers to invest in unprofitable risky ventures (the overinvestment problem).[130] And managers are more hesitant to invest in less risky yet profitable projects because the payoffs from these projects will accrue to debtholders in the insolvency states of the world (the underinvestment problem).[131] Reducing cash flow volatility and the probability of insolvency corrects both types of investment distortions.

Volatility in cash flow affects financing costs as well. It raises the probability of financial distress that would impose restructuring costs and disrupt relations with suppliers and customers. The higher risk of financial distress consequently increases the cost of debt capital and lowers leverage. Yet, debt financing has several advantages; most notably, interest payments are tax favored and debt may reduce managerial agency costs by compelling the disgorgement of free cash flow.[132] By increasing debt capacity, risk management restores the firm's ability to realize these gains. Volatility in cash flow also affects the availability of internal financing of new projects. If information asymmetries impede the raising of external capital, fluctuations in the available internal funds may prevent efficient capital budgeting.[133]

The goal of risk management, however, is not simply to reduce the firm's investment risk. After all, the cheapest way to manage risk might be simply for the firm to hold U.S. Treasury Bills. Diversifying operations reduces volatility, but there is considerable evidence that it distorts investment decisions by permitting inefficient cross-subsidization of projects.[134] A firm might also adapt to its operational risks by holding a buffer of cash or a line of credit, but this financial slack also may yield a temptation for managers to overinvest in unprofitable projects. Thus a firm may look instead to its capital structure and the design of its financial instruments to reduce the danger of insolvency. For example, it may shorten the ma-

130. Michael C. Jensen & William H. Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, 3 J. Fin. Econ. 305, 335–40 (1976).

131. Steward C. Myers, Determinants of Corporate Borrowing, 5 J. Fin. Econ. 147, 152–55 (1977).

132. Michael C. Jensen, Agency Cost of Free Cash Flow, Corporate Finance and Takeovers, 76 Am. Econ. Rev. 323, 324 (1986).

133. Kenneth Froot et al., A Framework for Risk Management, Harv. Bus. Rev., Nov.–Dec. 1994, at 91, 94.

134. E.g., Philip G. Berger & Eli Ofek, Diversification's Effect on Firm Value, 37 J. Fin. Econ. 39, 40 (1995) ("The potential costs of diversification include the use of increased discretionary resources to undertake value-decreasing investments, cross-subsidies that allow poor segments to drain resources from better-performing segments, and misalignment of incentives between central and divisional managers."); Hyun-Han Shin & René Stulz, Are Internal Capital Markets Efficient, 113 Q.J. Econ. 531, 533 (1998) (noting that diversification might lead to distortion in allocation of resources and creation of deadweight costs due to subsidization of losing divisions and expenditure of substantial resources on rent seeking and internal politics by divisional managers).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

turity of its outstanding debt or condition its interest obligations on the market prices of critical inputs or products. Flexible manufacturing techniques can also decrease the firm's vulnerability to market risks. For example, Hewlett Packard recently implemented a flexible assembly process that allows printers to be customized onsite by the customer rather than by the firm's principal manufacturing plants.[135] In short, risk management is a dynamic portfolio strategy that includes all of the firm's financial, insurance, derivative, and supply contracts.[136]

C. *Risk Management and Insurance in Thick- and Thin-Market Contracts*

Forward contracts and explicit options are perhaps the most widely discussed tools of risk management. It is common to think of risk management narrowly in the sense of transactions in traded derivatives (particularly in foreign exchange, interest rates, or commodities), where counterparties have a comparative advantage in hedging or diversifying the relevant risks. Yet, the market for derivatives is not complete and the bundle of risks facing a firm cannot be perfectly hedged. Insurance companies have stepped up their competition against financial instruments by offering industrial firms tailored insurance against bundles of risks, including earnings insurance.[137] In addition, a firm may exploit opportunities to outsource risk in contracts with its customers and suppliers.[138] As discussed below, embedded options play a key insurance role in these relationships.

1. *Forward Contracts and Option Contracts in Thick Markets.* — For ease of exposition, we adopt the conventional distinction between thick-market and thin-market contracts.[139] The paradigmatic thick-market con-

135. Alexander J. Triantis, Real Options and Corporate Risk Management, J. App. Corp. Fin., Summer 2000, at 64, 70.

136. See, e.g., Christopher Culp, Contingent Capital: Integrating Corporate Financing and Risk Management Decisions, J. App. Corp. Fin., Spring 2002, at 46.

137. Neil A. Doherty, Integrated Risk Management 523–36 (2000).

138. Recent developments in Financial Accounting Standards Board (FASB) standards have reflected the realization that such contracts, particularly if long term, have a significant impact on a firm's liquidity and risk profile, and should be disclosed in financial statements. The SEC now requires registrants to disclose the quantity, price, and duration of purchase obligations, SEC Filing Instructions, 17 C.F.R. § 229.303(a)(5)(i) (2004), pursuant to the directives of section 401(a) of the Sarbanes-Oxley Act (13(j) of the Securities Exchange Act). Sarbanes-Oxley Act of 2002 § 401(a), Pub. L. No. 107-204, 116 Stat. 745, 785 (to be codified at 15 U.S.C. § 78(m)).

139. The distinction between thick markets and thin markets is in some sense an artificial one. Most contracts fall between the two paradigms and are hybrids. Yet, it facilitates our analysis to adopt the artificial classification. As noted earlier, we define thick-market contracts as those contracts for which risk allocation rather than investment is the motivating purpose underlying the exchange. Thus, by thin-market contracts we mean to include all relational contracts where one or both parties contemplate making a relation-specific investment. The investments we have in mind would include the production of specialized goods, the development of human capital specific to a particular deal, or research to acquire information about future costs or prices.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

tract is traded in a liquid market at trivial transaction cost, and trading is motivated primarily by speculation or risk allocation. In a thick market, contracts are clearly divided between forward (or futures) contracts and option contracts. A forward contract is a simple purchase-and-sale contract in which the parties exchange exogenous risks, either because they have different expectations about future spot market prices or because they have different abilities to manage the market risk. Forward contracts in thick markets are not option contracts, and, unlike thin-market contracts, they do not embed options. Neither party has any meaningful termination or cancellation rights because of the symmetry between the value of a thick-market forward contract to the buyer and the performance cost of the seller. For the same reason, there is no real option of delay or abandonment for the parties to share. The substance of pure thick-market contracts renders inapt any discussion of compensatory damages: The enforcement of such forward contracts should be simply the specific enforcement of the market risk allocation.

When options exist in thick markets, they are explicit. Like forward contracts, they are motivated by heterogeneous information, expectations, or risk-bearing abilities. A seller may write an option in favor of the buyer if risk is costly to the buyer and the seller has a comparative advantage in risk management, such as the opportunity to diversify or to hedge the market risk.[140] The seller may offer a wide range of alternative option contracts to the buyer. Option contracts are defined by the underlying asset (e.g., the commodity), the exercise price, and the maturity date. For any given maturity and commodity (specifically, the probability distribution of the spot price of the commodity at maturity), the price of the option is a function of the exercise price. As before, we refer to this relationship by numerical pairings of option price and exercise price, (d, x), and note in the case of a call option that there is a range of such pairings, ranging from $(\hat{d}, 0)$ to $(0+, \hat{x})$, where $\hat{d}$ and $\hat{x}$ are the maximum values for the respective variables.[141]

The holder of an explicit call option contract may walk away from the purchase of the underlying commodity without making any payment other than the option price. The ex post loss incurred by the option writer does not affect the option price. Moreover, the option price is not

140. For example, the seller of a call option assumes the risk of the underlying asset value when it is out of the money; the buyer bears the risk while it is in the money. The parties might also agree to trade the call option if the seller (buyer) has private information indicating that the asset is overvalued (undervalued).

141. To our knowledge, there is no systematic explanation of the factors that determine which pairings are selected by options markets or become the most liquid. Commodities options tend to be issued near the money; the exercise prices are near the corresponding spot prices. It may be that options are too risky and difficult to value when the exercise price lies at the tail of the distribution of the underlying asset (i.e., are far out of the money). Thus, they may be correspondingly more difficult for the writer to hedge. Concerns about default and insolvency may reinforce the bias toward at-the-money options.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

intended to be even ex ante compensatory in the sense envisaged by the legal regulation of liquidated damages. When courts examine the reasonableness of ex ante estimates of compensation, they consider the seller's loss given the buyer's breach. But the seller is willing to write a call option as long as the option price compensates the seller for the cost of the option. This cost is a function of the entire distribution of payoffs, in all states of the world. The option price is simply the price at which the seller agreed to bear the market risk shifted by the option, given its exercise price and maturity. In a thick market, the option price is determined by the value of the option to the marginal buyer, which is equal to the cost of the option to the marginal seller. If the option price is negotiated between writer and holder (even though the spot market may be thick), the price will fall somewhere between the value of the option to the buyer and its cost to the seller. Thus, for the inframarginal option seller in the thick market and for the option seller in the negotiated deal, the option price will appear supercompensatory if a court considers whether it is a reasonable estimate of the ex post loss to the seller given breach. It is also supercompensatory in the sense that the seller may be overcompensated for its cost in producing any feature of a good or service delivered to the buyer. The penalty rule for liquidated damages does not, of course, apply to explicit option contracts. It does apply to bilateral contracts in thin markets, however, even if they contain embedded options that serve the same insurance purposes outlined above. Yet, even in thin markets, damages are effectively option prices that often have nothing to do with compensating the seller's ex post losses from breach.[142]

2. *Embedded Options in Thin-Market Contracts.* — As noted above, risk allocation is a principal purpose of thick-market contracts. Literal performance is often incidental and the parties may even choose to settle through monetary payment rather than physical delivery of the subject matter. Neither party has a termination right. Accordingly, there is a sharp distinction between forward and option contracts in thick markets. In contrast, a thin-market contract contemplates physical performance, and one or both parties may make investments specific to the contract that will be lost if performance does not occur. The widely recognized purpose of thin-market contracts is to protect the parties' specific investments. Yet, thin-market contracts also embed options in the form of termination rights in order to pursue the objective of efficient risk management. Thin-market contracts are incomplete and contracting parties are often superior insurers because of their private information and ability to control relevant risks. In these cases, the option price and exercise price

142. See Scott, Market Damages, supra note 1, at 1172–74 (concluding that "to the extent that market contracts are options on the future market, [they will], in general, better serve the contractual purposes of the broadest number of bargainers" because they "reliably measure the ex ante economic opportunities purchased by the contract" as opposed to the "injured party's ex post economic loss").

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

of the buyer's embedded call option are set to allocate optimally between the parties the risk of fluctuations in the buyer's valuation.

For example, consider two sets of contracts in the airline industry. First, we can express the ex post surplus yielded by a passenger flight in a highly stylized form as the difference between the value to the passenger and the cost to the airline. The passenger may avoid some of the lower valuations of the flight in a contract that permits her to breach and pay expectation damages. In the states of the world in which the passenger breaches, however, she still bears the risk of fluctuations in the airline's costs because they are reflected in the passenger's expectation damages liability. Even with the common law limits of foreseeability, mitigation, and uncertainty, the passenger would bear this risk despite having limited information and control. The parties may therefore contract for fixed liquidated damages. Moreover, as noted above, the passenger may be willing to pay a premium over the mean of the distribution of expectation damages. This liquidated damages amount is overcompensatory in that it exceeds reasonable estimates of the losses suffered by the airline when the passenger does not fly (e.g., the seat is sold to another passenger). This seems to explain why passengers pay penalties of $100 when they cancel their flights and their seats are resold at spot prices that are typically higher than advance-purchase fares. For example, on many routes the airplanes predictably fly at full capacity and yet the penalty is still assessed. Compensation is not the objective of the cancellation provision. The $100 reflects not only the expected cost to the airline, but also the value of the option to the passenger.[143]

Of course, some passengers want more insurance: They wish to transfer a larger portion of the risk associated with their valuation of the flight. These parties might choose a higher priced fare with a lower cancellation fee. This alternative is characterized in our analysis as a higher exercise price, x, and lower option price, d. The airline might well incur a loss from the passenger's decision to cancel, but our analysis emphasizes that the cancellation fee is the price paid for the call option on the

143. We find some appreciation of this argument in cases in which courts try to distinguish between liquidated damages, which are subject to the penalty rule, and provisions for alternative performance, which are not. As we noted above in our discussion of the Blockbuster litigation, when courts enforce alternative performance provisions, they recognize the value of options and that buyers therefore may be willing to pay a premium over the cost to the sellers of providing such options. See supra notes 11–13 and accompanying text. In *Pickens v. Blockbuster, Inc.*, customers claimed that the late fees were unenforceable liquidated damages because they were excessive: They were higher than the rental rate per diem. No. A102626, 2004 WL 339594, at *1 (Cal. Ct. App. Feb. 24, 2004). In affirming summary judgment for Blockbuster, the court found that consumers might rationally decide to hold on to the video and pay the late fee. Id. at *3. We argue that all termination rights should be viewed similarly: Liquidated damages are the price paid for the call option that a rational buyer might exercise in some states of the world. As in the case of any commodity or service, the price reflects the value of the option to the buyer and thus may exceed the cost incurred by the seller. The courts, therefore, ought not interfere with the parties' freedom to set these termination prices.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

flight rather than compensation for the airline's lost profit from termination.

The airline, in turn, passes some of the risk of fluctuations in passenger demand to its suppliers, including its aircraft manufacturers. For example, an aircraft manufacturer may be able to diversify the risk of fluctuating passenger demand better than the airlines if the manufacturer has a sufficiently diverse customer base. Moreover, the manufacturer also enjoys a comparative advantage in salvaging the aborted exchange by reconfiguring and reselling the planes.[144] Thus, the manufacturer may grant the airline an option to terminate during at least a portion of the manufacturing period.[145] The price for the option will be negotiated between the two parties and will lie somewhere between its value to the buyer and cost to the seller, depending on their relative bargaining power. Once again, the option price is unlikely to have a strong connection to a court's assessment of the manufacturer's foregone profit when the airline allows its option to expire without exercise.

The parties also may prefer that the seller use general manufacturing methods rather than cheaper specialized methods, so that the buyer can realize benefits from the option. The flexibility gains from general investments may thus outweigh the losses from foregone specific investment.[146] Specifically, the value to the holder of the option might exceed the cost of using general rather than specialized production methods. For example, Airbus manufactures an increasing number of different aircraft using the same technology platform and common parts, and thereby can sell valuable options to its airline customers to convert their orders into contracts for other aircraft.[147] The same flexibility could support an option

144. There are other examples of termination options that rely on the seller's comparative advantage in salvaging an inefficient exchange. In book publishing, for example, the consignment contract (providing the retailer a right of return) has evolved as the industry norm. Among other reasons for the free option to terminate, the publisher can reallocate inventory more efficiently because of specific knowledge about demand in other locations and because of economies of scale in coordination. Kandel, supra note 6, at 353–55.

145. The Marketing Director of Airbus Industrie, in noting the advantage of granting customers the option to terminate, indicated that his company was able to "double-book some delivery positions, on the basis that not all options were exercised" and that it seemed possible to "diversify some of this risk by double booking options from two airlines that were perceived to be countercyclical." Stonier, supra note 7, at 47.

146. Conventional contract theory examines the relative benefits of specific investment over general investment. More recent scholarship, however, recognizes that investment is specific or general to varying degrees. See, e.g., Eric Posner, Alexander Triantis, & George Triantis, Investing in Human Capital: The Efficiency of Covenants Not to Compete 13 (Jan. 2004) (unpublished manuscript, on file with the *Columbia Law Review*). A seller who uses general manufacturing techniques may be able to profit by selling the resulting flexibility in the form of embedded options. See Stonier & Triantis, supra note 7, at 164–65 (describing practices of aircraft manufacturing industry).

147. Airbus allows buyers to choose between the aircraft in its family of A318, A319, A320, and A321 with a very short lead time. Stonier & Triantis, supra note 7, at 160. These aircraft are built on the same production line and have many common components, so the

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

to terminate because the platform could more readily service other contracts.

In sum, thin-market contracts have multiple purposes. In addition to protecting relation-specific investments, thin-market contracts often serve an insurance or risk management function. Thus, parties to thin-market contracts will trade off the benefits of specific investment (which requires commitment) against the risk management benefits of general investment (which requires flexibility). The challenge of achieving the optimal balance further complicates the structuring of termination rights.

D. *Pricing Embedded Options in Thin-Market Contracts*

The pricing of embedded call options is determined by the cost to the seller of providing flexibility and its value to the buyer. It is likely to be quite heterogeneous across circumstances. This observation motivates our argument in Part III against a damages default rule that purports to set a "price" for breaching and the penalty doctrine that constrains the parties' freedom to do so. The pricing of termination options is itself an interesting avenue for analysis and, in the following section, we offer a preliminary outline of the relevant factors that determine the option price.

1. *Payment and Possession During the Option Term.* — When an explicit call option is written in a thick market (e.g., a commodities option), the option holder (the buyer) pays the option price, but does not advance the exercise price until she exercises the option. This is not always the case with the embedded call options we have described. In some cases, the buyer pays the entire contract price that includes both the option price and the exercise price, and is entitled to a refund of the exercise price when either she surrenders her option or it expires. In other cases, the buyer pays nothing up front and simply incurs the contingent liability imposed by a court when she walks away from the contemplated exchange. Or the buyer may leave a nonrefundable deposit equal to a fraction of the contract price or may agree to stipulated damages. A potentially significant difference among these possibilities is the extent to which the seller receives money to hold and invest during the option term. If the seller holds some of the exercise price, the buyer foregoes interest on this sum and must recover this lost income in the option price. For example, a money-back guarantee is not a free call option since the buyer gives up the income on the refundable portion during the option term.

---

decision as to which aircraft is built can be delayed. Id. at 160 n.2. Airbus explicitly markets contract options and offers guidance to buyers in the valuation of these terms. Stonier, supra note 7, at 48–49. "This facilitated [Airbus's] ability to allow customers to defer, until as late as possible, the decision as to which size of aircraft within the family they would actually take delivery of." Id. at 41.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

A holder of an explicit call option typically does not enjoy possession of the underlying asset, and thereby its value in use, until she exercises her option to acquire the asset. Recall that the value of an option comes from the resolution of uncertainty through the production of new information during the option term. For many thin-market goods whose value is subjective rather than purely market driven, the seller will be inclined to give possession and use of the good to the buyer in order to encourage information production. Rent-to-own centers capitalize on this phenomenon in making sales, and explicitly lease the good to the prospective buyer during the term of the option. Yet, it is important to bear in mind that parties who let their options expire owe the duties of a bailee to return the good in reasonable shape. And, in some cases, the use value of the good during the option period may roughly offset the seller's income from the buyer's up-front payment.

2. *Exogenous Risks, Moral Hazard, and Adverse Selection.* — Parties write embedded options in part to exploit comparative advantages in bearing exogenous risks. As with any service, insurance will be provided if the risk-bearing cost to the provider is lower than the value to the insured. Thus, if an embedded call option simply entailed the transfer of exogenous risk from buyer to seller, the seller would write the option that maximized the surplus created by the difference between the option's value to the buyer and its cost to the seller. The seller would agree to assume the range of risk over which it has a comparative risk-bearing advantage. As noted earlier in Part II.B, these measures are functions of the multivariable risk profiles of each party. Given that the underlying exchanges occur in thin markets, the option price is the product of bargaining between the seller and buyer, who will divide the surplus. Thus, it is likely that the price the buyer pays for the option will be higher than the cost of the option to the seller. The seller's cost will be a function of the volatility of the buyer's valuation, as well as the length of the option term.

Given the close connection between options and insurance, it should not be surprising that the determinants of the structure of embedded options should also invoke familiar considerations about moral hazard and adverse selection. The parties cannot completely control for these information problems because the courts cannot verify ex post the cause of any realized loss. In the face of such concerns, the parties must weigh the various factors in choosing whether to include an embedded option and, if so, its exercise price and option price. In the traditional insurance contract, the relevant private information and hidden action lies with the insured party. In the context of an embedded option in a sales contract, however, there are concerns about adverse selection and moral hazard on both sides.

3. *Seller's Private Information and Control.* — The seller may provide insurance to the buyer in the form of a call option if the seller has private information about the good that affects the buyer's valuation. A simple contract would face a significant adverse selection problem if the range of

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

possible quality was large. The buyer might offer a price associated with the average value of the product, but this may drive from the market the high-quality sellers, leading the buyer to further discount the price. Warranties are well known as signaling devices that bridge the information asymmetry between seller and buyer.[148] Their enforcement, however, requires a court to determine whether the warranty has been breached. The costs of verifying breach may be large relative to the value of the signal, either because the evidence is indeterminate or because the warranty standard is opaque.[149] Therefore, the seller may give the buyer instead the right to return the good or cancel the deal without the burden of proving breach or incurring prospective litigation costs.[150]

The seller's private information is the most common explanation for the return policy of retailers of goods, particularly those whose value may not be transparent. If the only source of volatility in the buyer's valuation was the quality of the good, the seller could ensure that no buyer would wish to exercise her right of return. The seller thus would be prepared to issue the option for free.[151] The buyer's valuation of the good is rarely due entirely to factors within the seller's private knowledge, however. To the extent that there are exogenous risks, such as technological innovations during the option term that affect the buyer's valuation, the seller bears a cost in writing the option. She may therefore be inclined to impose a fee. In some cases, however, she may bear the cost of the option as a signal of her favorable private information. Thus, the right to return effectively falls within a category of gifts that sellers give to market their products to prospective buyers. Instead of giving the prospective buyer a free dinner, a round of golf, or a tote bag, the seller gives a free option.

When the exogenous risk increases, the seller may change her strategy to either begin charging an option price or not offering the option at

148. See generally George Akerlof, The Market for "Lemons": Quality Uncertainty and the Market Mechanism, 84 Q.J. Econ. 488, 499–500 (1970).

149. In this respect, we note the vague default standards for warranty liability under the UCC. Under the Code's implied warranty of merchantability, for example, the buyer, in order to establish a breach of warranty, must show that the good is not "fit for the ordinary purposes for which such goods are used." U.C.C. § 2-314(2)(c) (2003).

150. See Katz, supra note 1 (manuscript at 43) (explaining that "a warranty is itself a kind of option—namely, an option to return the goods [no questions asked] if they do not measure up to the promised quality").

In some cases where the seller has private information, the buyer may be able to acquire the information at some cost. But if the investigation requires an investment that the buyer will not recover if she fails to purchase the product, then the seller might hold up the buyer during subsequent negotiations over price. There are several ways in which the buyer's investment in research may be protected. One way is for the seller to give the buyer an option on the good at a specified exercise price.

151. For example, sellers of electronics equipment have significant private information relative to their customers concerning the capabilities of the products they sell. Other examples of sellers' private information where a right of return is common include oriental rugs, works of art—such as paintings and sculptures—and similar goods whose value is based on a combination of provenance and aesthetic qualities.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

all. For example, hotels and resorts are more likely to levy cancellation charges in high season than in low season even though the likelihood of finding a substitute customer is greater in the former case.[152] Our analysis suggests that the cancellation charge is not compensation for lost revenue, but rather the fee for the option held by the customer. The call option is more costly to write (and probably also more valuable to the buyer) in high season because customer demand tends to be more volatile than in low season where the truncation of the distribution at zero is more significant.[153] In turn, in low season the volatility in the value of the reservation is more likely to be the product of the seller's private information. For example, the seller will have private information concerning the quality and range of services available during the off season.

Suppose that the value of the good to the buyer depends on effort or investment by the seller that is not contractible. In other words, the buyer's valuation is a function of the seller's cooperative investment. For example, assume that the buyer commissions a portrait of herself and her children by a well-known artist whose prior work demonstrates great skill in capturing the essence of his subjects. The artist might be induced to make the efficient level of investment by giving the buyer a call option with an exercise price equal to the value that results when the seller exerts the efficient level of effort. As noted earlier, however, this solution is undone when the buyer can force the renegotiation of the exercise price and can thereby hold up the seller with respect to her specific cooperative investment.[154] If the seller could precommit not to renegotiate, then the option serves the purpose of inducing the efficient investment or effort by the seller. Retailers, for example, have largely been successful in preventing renegotiation by constraining the authority of their sales agents to modify the price of goods offered for sale.

---

152. See Katz, supra note 1 (manuscript at 20) (citing anecdotal evidence that suggests "when demand is high and rate[s] are marked up, the innkeeper charges a stricter deposit—and not just a larger amount to correspond to the higher in-season room rate").

153. To illustrate, consider a ski resort with 100 rooms. On average, it fills 75 rooms in high season and 2 rooms in low season. Compare the effect of an adverse event such as bad weather on the demand in both seasons. In the high season, the resort could lose up to 75 paying customers; in the low season, its losses are bounded at 2 customers, regardless of the severity of the adverse contingency.

One of our students observed that inspection clauses in residential home purchase agreements in Washington, D.C., became tighter or disappeared when that market heated up several years ago. D.J. Moore, Understanding Real Estate Hot Markets 8 (Dec. 2002) (unpublished student paper, University of Virginia School of Law) (on file with the *Columbia Law Review*). We speculate that the cause was the higher volatility rather than heightened demand. To see why volatility matters rather than absolute price levels, consider a hot market in which the prices have jumped but are expected to remain constant for the next year. Under these circumstances, we would predict that the seller's cost of writing the option is relatively small and the seller may do so to signal its private information about the quality of the property.

154. See supra notes 96–98 and accompanying text.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

4. *Buyer's Private Information and Control.* — We now turn to consider the buyer's private information and control that has the same effect as adverse selection and moral hazard in insurance. If buyers are heterogeneous in the volatility of their valuations, adverse selection concerns may lead sellers not to write costly call options. For example, a buyer from a retailer with a policy of free returns may underinvest in search activity to determine the suitability of the product to her tastes. When this buyer is an agent of her family, she may fail to invest efficiently in learning whether the product suits the tastes of other family members.[155] In general, suppose that buyers are fickle to varying degrees, but the seller is unable to discriminate among buyers. The seller would charge an option price that reflected the average sensitivity of its buyers. The less fickle buyers, therefore, would cross-subsidize the options of the more fickle buyers and the less fickle might exit as a result. A competitor retailer (an entrant) may see an opportunity to deny its customers the right of return in order to attract with lower prices the business of the less fickle buyers, thereby leaving the original retailer with a more fickle pool and higher costs that must be recovered by raising prices.

There are many possible equilibria, depending on conditions such as the risk aversion of consumers in the market and the volatility of their tastes over time. The entrant might draw away all buyers, leaving an industry without return options. Or, there may be a separation of the market in which the competitors coexist; for example, some restaurants take reservations while their competitors do not. Finally, the original firm may be able to hold on to most of the buyers because of its comparative ability to bear the buyers' valuation risks or because of the presence of significant seller-side private information or control. For example, the seller might also offer different combinations of option price and exercise price in order to discriminate among buyers with heterogeneous valuations.[156]

The insurance written by the seller who gives an option may also give rise to moral hazard in the buyer's actions, both during and before the contract period. Of course, where the buyer has possession of the good during the option term, she will not have the efficient incentive to take care of it. Moral hazard also comes in more subtle forms. During the contract, the buyer who is protected by insurance will be motivated to forego measures that dampen (or take actions that increase) the volatility

155. This observation is distinct from Richard Craswell's theory of the effect of damages on search. His focus is on a party's precontract search for the appropriate contracting partner. See Richard Craswell, Offer, Acceptance, and Efficient Reliance, 48 Stan. L. Rev. 481, 500 (1996) ("The damage remedy may also affect the extensiveness of the search for contracting partners, and the choice from among the available partners." (footnote omitted)). We are focusing on the buyer's effort in estimating the value of the good itself—e.g., whether the family really needs the television.

156. A different price discrimination model, which focuses on variability in buyer valuation of the contract exchange itself, is presented in Barry Adler & Alan Schwartz, The Lost Volume Problem 4–5 (Aug. 17, 2004) (unpublished manuscript, on file with the *Columbia Law Review*).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

of her valuation of the product in order to maximize ex post the value of her option.[157] Because sellers cannot control for buyer care in granting return options, buyers have the incentive to underinvest in actions that will reduce volatility. Whether the information obstacle is one of adverse selection or moral hazard, the seller might require the buyer to bear some portion of the cost of returns by requiring the expenditure of time in exercising the option even though it is nominally free. Moreover, some returns are limited to "store credit" that further constrains the buyer's choice and thus motivates the buyer to internalize some of the cost of exercising the option.

5. *The Bargaining Environment and Heterogeneity.* — We have established that, as a call option, a termination right in a goods or services contract entails a sale of insurance from the seller to the buyer. Normally, the seller would not offer insurance for a price below its cost and the buyer would not agree to pay a price higher than her valuation. Assuming that the call option is efficient, the option price will be between the seller's cost and the buyer's valuation. Where it lies in this range depends on the contracting environment. In simple terms, a competitive market will force the option price down to the seller's cost. Anticipating our criticism of the penalty rule, we emphasize that the seller's cost of writing the option is equal to neither its realized loss from breach (the non-exercise of the call) nor even the expected loss conditioned on breach. Outside the perfectly competitive environment, the seller can expect to recover a profit above her cost. For example, a discriminating monopolist may be able to extract all the option value from the buyer. And, in an individually negotiated bargain, the portion of the surplus enjoyed by the seller will depend on the parties' relative bargaining power.

In sum, the embedded options that are common in commercial and consumer contracts are a function of trade-offs arising from the interaction of comparative advantages in bearing exogenous risks and the problems of hidden action and hidden information between sellers and buyers. Parties are more likely to agree to embedded options at lower option prices where the seller has superior ability to bear exogenous risks and enjoys private information about, and relatively more control over, the contract good. Of course, in designing optimal options, parties must weigh these factors along with the efficient breach and efficient investment concerns that are currently the preoccupation of most academic work in the economic theory of contracts.[158] Given the heterogeneity of sellers and buyers and of the subject matter of exchanges, the prices of termination rights or call options vary widely. As we have discussed, these prices can be expected to deviate greatly from the compensation that matters in contract damages: the amount necessary to put the seller in

157. Triantis, Insolvency and Bankruptcy, supra note 1, at 685.

158. See supra Part I.C.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

the position she would have been in if the exchange were completed. The seller's cost of writing the option is not equal to this amount and, in many bargains, the seller can extract some of the value of the termination right to the buyer. In Part III, we turn to a consideration of the ways in which these important insights inform our understanding of optimal damages default rules.

III. Optimal Default Remedies

As we discussed in Part I.B, courts and legislatures have firmly entrenched the compensation goal in a majoritarian default of full expectation damages.[159] The theory of embedded options advanced in this Article suggests, however, that optimal contract damages have little to do with compensation and that they are as context dependent as the contract price of goods or services. This heterogeneity calls into question the wisdom of having a majoritarian default, particularly one that embraces the compensation principle. Specifically, the expectation damages default has two fundamental drawbacks. First, it is not the optimal choice for a majority of contracting parties. Second, the compensation principle impedes the freedom of parties to contract for alternative damages or termination rights. The impediment takes an extreme form in the penalty rule prohibition of supercompensatory liquidated damages. We have shown that when the seller can extract some of the value of an embedded option to the buyer, the option price (or damages) might appear to be supercompensatory. By striking down such a provision, the penalty rule acts as an undesirable form of price regulation. More generally, the compensation principle has achieved a salience such that it engenders a judi-

159. Although various authors have questioned the merits of the expectancy default, no one thus far has doubted the case for a majoritarian default damages provision or the salience of the compensation principle in shaping the preferred default rule. See, e.g., Charles J. Goetz & Robert E. Scott, Enforcing Promises: An Examination of the Basis of Contract, 89 Yale L.J. 1261, 1281–86 (1980) (determining optimal damages formula based on reliance interest). Recently, Richard Craswell summarized the difficulty in selecting an economically efficient default in his critique of the categories of compensatory damages advanced by Fuller and Perdue. After reviewing the various decisions made by contracting actors (search, precautions, reliance, and trade), he suggests that

> [t]here is . . . no reason to suppose that the totality of economic effects will *always* favor an award of expectation damages. . . . [W]henever different amounts of damages would be optimal for each of the different incentives to be optimized, the measure that is optimal when *all* of the relevant incentives are considered will often be some hybrid or intermediate number [of the expectation, reliance, and restitution interest measures].

Craswell, Against Fuller and Perdue, supra note 18, at 109–10.

Craswell nevertheless believes that any useful framework for analyzing remedies would have some significant link to compensation. See id. at 156–58 ("[T]he expectation remedy has a number of effects (perhaps more than any other single remedy) that are relevant from a normative standpoint [and] make[ ] that remedy a convenient baseline for policy analysis."). He comes to a similar conclusion in his critique of Fuller and Perdue's three-way classification of expectation, reliance, and restitution interests. Id.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

cial bias against provisions that depart from the compensatory damages default.

We argue in this Part that it is not sufficient to admonish the courts to specifically enforce all option prices, whatever their form. In place of the "majoritarian" expectation default term, we propose a radical departure in favor of specific performance of exchanges contemplated in commercial contracts in order to encourage parties to specify explicitly their termination rights (including any damages term) at the time of contracting. In light of the fact that merchant sellers typically draft consumer contracts, we argue that consumer buyers should hold free options unless the contract provides otherwise.

A. *Conditions for Efficient Majoritarian Defaults*

The case for majoritarian default rules in contracts rests on the premise that state institutions, such as courts and legislatures, sometimes can design contract provisions at lower cost than the parties could themselves.[160] One source of this advantage may be economies of scale. Private parties who bear the cost of creating efficient terms to address contracting problems cannot capture the full social gain because they cannot collect a fee from later parties who might copy those terms.[161] As a result, contracting parties may underinvest in contract design. The efficiency of default provisions relies on the satisfaction of two key conditions. First, parties must have the freedom to contract for their own terms because agreements occur in heterogeneous circumstances and the state cannot feasibly write contracts that suit all environments. Second,

160. A "majoritarian" default rule purports to provide a contract term that the broadest number of bargainers would have stipulated for themselves had their contracting costs been lower. These defaults are justified, therefore, to the extent that they are widely suitable to many contracting parties, thus eliminating both the resource cost and the error cost of negotiating over the proposed term. Robert E. Scott, A Relational Theory of Default Rules for Commercial Contracts, 19 J. Legal Stud. 597, 608–09 (1990). On the other hand, a "bargain-forcing" default is one that encourages the parties to bargain explicitly over the term in question by penalizing one party (or both) should they remain silent. These rules are not set to reflect the term preferred by most bargainers. Bargain-forcing defaults share a familial resemblance with so-called "information-forcing" default rules that induce one party to share important information with the other. Id. at 609–10. We use the bargain-forcing characterization in this paper because it embraces related contract defaults, such as the indefiniteness doctrine, that encourage the parties to bargain over heterogeneous factors, such as contract price, at the penalty of having the agreement held unenforceable. For discussion, see Robert E. Scott, A Theory of Self-Enforcing Indefinite Agreements, 103 Colum. L. Rev. 1641, 1645–59 (2003) [hereinafter Scott, Self-Enforcing Indefinite Agreements].

161. For analyses of how legal practice innovations spread, see Gerald F. Davis, Agents Without Principles? The Spread of the Poison Pill through the Intercorporate Network, 36 Admin. Sci. Q. 583 (1991); Gerald F. Davis & Henrich R. Greve, Corporate Elite Networks and Governance Changes in the 1980s, 103 Am. J. Soc. 1 (1997); Charles J. Goetz & Robert E. Scott, The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms, 73 Cal. L. Rev. 261, 291–93 (1985) [hereinafter Goetz & Scott, Limits of Expanded Choice].

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

although the state can better internalize the benefits of default provisions to all parties, the economies of scale available to the state depend on the degree to which the needs of contracting parties are homogeneous and observable by the state.[162]

The first condition of an efficient default—the ability of parties to contract for their own, different term—is generally satisfied in contract law. As we elaborate below, however, this condition is violated in the case of damages for breach because of the dominance of the compensation principle that supports the penalty doctrine and a more general judicial bias against enforcing attempts to contract away from the compensation default. The second condition—the state's cost-efficiency advantage in designing defaults—is difficult to satisfy with respect to contract terms that are highly context specific. In light of the heterogeneity and complexity of commercial and consumer contracts, the economies of scale in fashioning default terms are generally outweighed by the informational advantage enjoyed by the parties themselves. Simple default rules are often inadequate in a complex world.[163] We have argued that this is the case with respect to the design of optimal contract damages. It is tempting to prefer standards rather than rules to address this type of heterogeneity. Indeed, a substantial set of default standards addressing a wide range of contracting terms and contexts have been proposed in Article 2 of the UCC and in the Restatement. But as one of us has demonstrated elsewhere, standards give rise to a variety of undesirable effects and thus should be avoided as a matter of contract policy.[164]

---

162. For discussion, see Schwartz & Scott, supra note 97, at 598–601 ("There are three criteria for a good default rule: It must be conditioned on only a few possible states of the world, be relatively simple in form, and be efficient for a wide variety of contract parties.").

163. Schwartz and Scott have argued that an efficient default rule is one that is relatively simple in form and suitable for a wide variety of contracting parties. Id. The criterion of simplicity is a function of institutional competence. Id. at 598. "Courts cannot conduct investigations into the efficiency properties of [all] possible rules and rule combinations." Id. The suitability criterion is even more difficult to satisfy in connection with breach remedies because, as we have shown, contracting parties are exceptionally heterogeneous regarding termination provisions. See supra Part I.D. State-created default rules may not even be feasible where there is asymmetric information. Defaults must be conditioned on information that the enforcing authority is able to observe. Schwartz & Scott, supra note 97, at 605–08. A default rule that is conditioned on unverifiable information would create a moral hazard. Id. Consequently, parties will routinely contract out of these possible defaults.

164. Schwartz and Scott argue that standards increase the uncertainty of litigation, reduce the likelihood of settlement, and encourage parties to increase their expenditures on litigation. See Schwartz & Scott, supra note 97, at 601–03. Richard Craswell has shown that courts have over time effectively created a default standard based on the broad categories of overexpectancy, expectancy, or underexpectancy. See Craswell, Against Fuller and Perdue, supra note 18, at 138–54 (suggesting that cases fit into framework based on these three categories).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

B. *Optimal Damages Default in Thick-Market Contracts*

Why do parties enter into fixed-price contracts for future delivery of goods that are traded in a thick market? After all, one can always acquire the goods on the spot market at the prevailing price without negotiating a forward contract. In the pure thick-market contract, there is no specific investment, no efficient breach, no information asymmetry, and no moral hazard. As we discussed in Part II.C, parties typically enter fixed-price contracts in these markets either to speculate on private information or to shift market risks. In the most liquid markets, these parties often do not contemplate physical delivery, but only a closing of their respective positions through the payment of money. In this light, there is no reason for the court to take any action other than to enforce the risk allocation by awarding market damages.[165]

Consider a contract to deliver at a future date 1,000 tons of a stipulated commodity at a contract price of $200,000. Suppose that the seller later contracts to fill the order from a third-party supplier for $150,000. The buyer then breaches and, at the time of performance, the market price of the commodity is $100,000. The third-party supplier subsequently releases the seller from his contract. The market measure of damages is the $100,000 difference between the market and contract price. The buyer argues, however, that the seller's loss is only $50,000 in foregone profits and that this compensatory amount is the appropriate measure of damages.[166]

Before we determine the optimal default rule to govern the buyer's breach, we must first understand the motivations that lead parties to write this contract. If the buyer only wanted to guarantee a supply of the commodity at the date of delivery, the parties could have contracted to pay the market price on that date. The fixed-price contract, however, has effectively shifted the risk of price fluctuations. After the contract, the seller bears the risk of price increases and, in turn, has purchased the reciprocal opportunity to hold the buyer to the risk of a price decline. The seller can deal with this contract risk in a number of ways. The seller may choose to self-insure and bear the entire risk internally. If so, the seller may wait and purchase the contract goods on the spot market before the contract delivery date. Alternatively, the seller can lay off the risk, as in our example, by entering into a forward or option contract to acquire the goods from a third party at a fixed price.[167] In the example,

165. The following analysis of optimal damages measures in thick-market contracts draws on the text in Scott, Market Damages, supra note 1, at 1160–79.

166. The facts of this hypothetical are loosely drawn from Nobs Chemical v. Koppers Co., 616 F.2d 212, 214 (5th Cir. 1980).

167. Typically, the seller is a middleman who has entered into a purchase contract with one party and a resale contract with another, thus eliminating its price risk. Goldberg, Framing Contract, supra note 106 (manuscript at pt. 4.3). The problem arises when, as in the hypothetical, the buyer breaches but the first seller fails to pursue his claim against the middleman. But these are two separate contracts and, under basic privity of contract

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

the seller purchased a call option from the third party and presumably paid something for it. If the court imposed on the buyer a liability only equal to $50,000 in foregone profit rather than the $100,000 market damages, it would permit the buyer to appropriate (without paying) the benefit of the seller's contract with the third party without compensating the seller. This policy will deter such third-party contracts in the future and, as a result, perhaps even similar initial contracts between sellers and buyers. The courts, therefore, should not read options (e.g., to terminate and pay foregone profits) into forward contracts. Forward contracts entail clear risk allocation that is enforced by granting the seller damages based on the full $100,000 market risk.

Driven by the compensation principle, however, most courts have declined to award market damages in similar contexts.[168] An especially attractive authority has been section 1-305 of the UCC, which expressly incorporates the compensation principle.[169] This provision has led courts to conclude that the market damages measure is not appropriate where the market fluctuation deviates substantially from the ex post economic loss to the promisee.[170] But these decisions are based on a misunderstanding of the motivation for thick-market contracts, which is simply the allocation of risk. The use of the compensation principle to depart from market damages in thick-market contracts is inconsistent with the parties' risk management goals. In these thick-market contexts, the courts should simply enforce the parties' risk allocation. If parties prefer an alternative risk-shifting mechanism offered by option contracts, the court should similarly simply enforce the risk allocation for which they have bargained.

C. *Optimal Damages Defaults in Thin-Market Contracts*

Most contracting does not occur in the paradigmatic thick-market context. The issue of optimal damages rules is more complex when we

---

principles, the middleman's supply contract with the first seller should be irrelevant to the breaching buyer. Id.

168. See, e.g., H-W-H Cattle Co. v. Schroeder, 767 F.2d 437, 439–40 (8th Cir. 1985) (affirming that buyer who had already voluntarily limited its market price is not entitled to market damages); *Nobs*, 616 F.2d at 215 (holding that seller is entitled to lost profits but not market damages, when market damages are substantially greater); Coast Trading Co. v. Cudahy Co., 592 F.2d 1074, 1083 (9th Cir. 1979) (finding that seller is entitled to actual losses when market damages would provide windfall); Union Carbide Corp. v. Consumers Power Co., 636 F. Supp. 1498, 1503 (E.D. Mich. 1986) (holding that seller receives only lost profits damages where market price damages would overcompensate seller); Allied Canners & Packers, Inc. v. Victor Packing Co., 162 Cal. App. 3d 905, 915 (1984) (limiting award of damages to actual loss rather than market damages).

169. "The remedies provided by [the Uniform Commercial Code] must be liberally administered to the end that the aggrieved party may be put in as good a position as if the other party had fully performed . . . ." U.C.C. § 1-305(a) (2003) (formerly § 1-106).

170. But see Trans World Metals, Inc. v. Southwire Co., 769 F.2d 902, 907–08 (2d Cir. 1985) (finding that compensation principle does not trump market damages simply because contract-market differential exceeds promisee's economic loss); Tongish v. Thomas, 840 P.2d 471, 476 (Kan. 1992) (same).

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

turn to thin-market contracts. Thin-market bargains (by which we include also hybrid thick-thin markets) are motivated by a wider range of objectives than thick-market contracts. Scholars appreciate that a common goal is the protection of specific investments and that, although expectation damages do promote efficient exchanges, they induce excessive specific investment. This inefficiency of expectation damages is compounded by the observation made in this Article that parties structure termination rights, including damages, to pursue risk management goals. A central insight of the embedded options approach is that there are a wide variety of termination provisions and that the measure of damages for breach is merely a subset of this larger family of termination options. Like other termination fees, contract damages are essentially the nonrefundable portion of the contract price. And, as we have demonstrated, we would normally expect parties pursuing risk management efficiency to set the price of the option somewhere between the option's value to the buyer and its cost to the seller.

This insight has been missed by courts and commentators who have been wedded to the compensation objective, and the oversight has contributed to various inefficient doctrines and misguided analysis. We first discuss the most glaring among these, the penalty doctrine. We then demonstrate that the compensation objective leads the courts into debates over the appropriate measure of expectation damages that are frequently beside the point. Finally, we suggest that there is a pervasive bias among judges and even lawyers that impedes the full exploitation of risk management opportunities through termination rights that have nothing to do with the compensation of sellers for ex post losses caused by the buyer's breach.

1. *The Penalty Rule.* — As noted in Part I.B, a liquidated damages provision is void as a penalty unless the damages are "reasonable in the light of the anticipated or actual harm caused by the breach, the difficulties of proof of loss, and the inconvenience or nonfeasibility of otherwise obtaining an adequate remedy."[171]

Once we see the damages award as essentially a price for a service (insurance against fluctuations in the buyer's valuation), the penalty doctrine becomes a major impediment to efficient contracting. Contracts scholars have previously demonstrated that the premise underlying the regulation of liquidated damages—that parties have no good reason to write damages clauses that depart from the compensation principle—fails for several reasons.[172] Options analysis supplies an even more devastat-

171. U.C.C. § 2-718(1); accord Restatement (Second) of Contracts § 356 (1981).

172. Some of the criticism of the penalty rule has come from within the compensation norm. For example, as a result of the doctrines of foreseeability and uncertainty, expectation damages are limited to the losses from breach that the promisee expects to be able to verify to a court. A liquidated damages clause that incorporates observable but nonverifiable values thus will be vulnerable to a penalty claim even when the clause accurately measures the promisee's lost expectation. As a consequence, sophisticated

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

ing critique of the penalty rule and, generally, of the doctrinal mandate that liquidated damages adhere to the compensation principle. As we have emphasized, damages are merely termination fees and thereby call option prices. If parties use termination rights to allocate risks, those fees or prices do not purport to estimate the seller's loss resulting from the buyer's termination. This is patently true with respect to the courts' focus on ex post loss and on ex ante reasonable estimates of loss given breach. Indeed, the sale of embedded contract options will often yield a profit to the seller above the cost of providing insurance to the buyer, in much the same way that the seller of a car will make a profit on the sale of options such as a sunroof and CD player. The option fee is a function of the value of the option to the buyer and not a function of the ex post loss suffered by the seller. Thus, there is no penalty in any meaningful sense of the word.

Price regulation is not the domain of contract law. There is no reason why a seller should not be able to profit on providing insurance to her buyers.[173] Given the heterogeneity of call options and contexts, as well as their complexity, courts would be far less suited to judge the fairness of option prices than the prices of the sunroof and CD players mentioned above. Indeed, the value of a call option to a buyer is a function of the variance of the buyer's valuation, which is significantly more difficult to assess than mean values. It seems better policy to permit parties to bargain explicitly over call option prices.[174]

2. *Doctrinal Debates over the Measurement of Expectation.* — A number of doctrinal controversies neglect the important contract goal of risk management. The measure of expectation damages is a point of controversy in lost volume sales cases. In the paradigmatic lost volume case, the seller

---

parties are discouraged from using liquidated damages clauses even when these clauses would otherwise be optimal. See Goetz & Scott, Liquidated Damages, supra note 20, at 578–83 (demonstrating that parties would rationally use liquidated damages clauses to insure a contractual performance that was nonverifiable and that courts view such clauses as supercompensatory and hence unenforceable). Other authors suggest that supercompensatory damages may be necessary to achieve efficient relation-specific investment. See supra note 103.

173. Ian Ayres and Barry Nalebuff discuss an unfair trade practice action brought by the state of Connecticut against a car rental company that charged a fee of $150 if the rented car was driven over eighty miles per hour for more than two minutes. They note that the state did not challenge a provision in the rental contracts charging $5 per gallon of fuel if renters fail to return a full gas tank. They suggest that courts may view terms that are framed explicitly in terms of price (such as the $5 per gallon of gas) more charitably than they view terms that are framed as "penalties." Ian Ayres & Barry Nalebuff, Connecticut Speeder-Friendly Crackdown, N.Y. Times, Aug. 31, 2001, at A19.

174. In this vein, however, there may be a relatively weak vestige of a case against supercompensatory liquidated damages. To some buyers, notably consumers, the term "damages" may suggest that the seller represents his loss from breach as equivalent to that amount. If this amount were supercompensatory, the buyer might be induced to overestimate the cost of the option to the seller and thereby overpay for it. At most, this argument speaks in favor of striking down penalties only when explicitly framed as liquidated damages.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

argues that the buyer's termination deprives the seller of a sale, so that the seller's economic loss systematically exceeds market damages.[175] In these cases, scholarly debate has focused on how much of the seller's selling costs or overhead were "consumed" by the breaching buyer and whether the default measure of damages ought to be the full profit lost by the seller (which may be overcompensatory) or incidental damages (which may be undercompensatory). But the focus on lost volume and selling costs is a red herring. Rather, the choice between market damages and lost profits from a foregone sale is a choice between alternative termination provisions or embedded call options.

The lost volume debate thus ignores the full range of alternatives. As described in Part II, the choice of termination provisions is far richer than the debate concedes: The parties may choose among pairs of option prices and exercise prices, (d, x). The parties do not intend that the option price reflect the seller's loss from the terminated sale. As we have discussed, a contract may have an option price that yields a net profit to the seller from writing the option. In other contracts, the seller might offer the call option at below its ex ante cost, as a signal of quality or as a loss-leading marketing strategy. In sum, whether a given volume seller would have chosen to write an option to a buyer and the price at which the option would be offered simply cannot be determined a priori. In any event, as long as risk management is a contracting goal, the termination provision is unlikely to have much to do with compensating the seller for the ex post loss of a sale.

The compensation principle (and the penalty rule) corrupts the way courts apply lost volume damages in cases in which the seller has required the buyer to pay a nonrefundable deposit. The express terms of section 2-718(2) of the UCC effectively limit the deposit that can be retained by the seller to either $500 or the amount permissible as liquidated damages under the penalty rule.[176] In response to this regulation of deposits,

---

175. See, e.g., Neri v. Retail Marine Corp., 285 N.E.2d 311, 314–15 (N.Y. 1972) (asserting that seller should be able to recoup lost sales volume from buyer's breach). The lost volume claim is particularly salient when the market damages measure (the difference between the contract price and market price) is zero. Sellers argue that, although they did in fact resell the contract goods at the contract price, the second buyer would have purchased anyway. Therefore, if the breaching buyer had performed the contract, the seller would have realized two profits rather than just one. Thus, the seller claims the lost profit from the breached contract as compensation for his lost expectancy. Scott & Kraus, supra note 19, at 1100–03.

176. The UCC states:

> Where the seller justifiably withholds delivery of goods because of the buyer's breach, the buyer is entitled to restitution of any amount by which the sum of his payments exceeds
>
> (a) the amount to which the seller is entitled by virtue of terms liquidating the seller's damages in accordance with [the penalty rule], or
>
> (b) in the absence of such terms, twenty percent of the value of the total performance for which the buyer is obligated under the contract or $500, whichever is smaller.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

many sellers sue for lost volume damages.[177] While scholars have criticized the courts for granting lost volume damages in this circumstance, the impulse of the courts may well be to provide the seller a "make up" for the invalidation of the contracted-for termination fee.[178] In turn, the precedential effect of these decisions may lead courts to grant lost volume damages in other contexts, where it is likely that the parties would have preferred a low-cost option.

The compensation principle has given rise to a number of other doctrinal controversies that similarly neglect the contract goals of risk management. Consider, for example, the case in which a buyer (or lessee) purchases the right to mine the seller-lessor's land in return for a royalty on extracted ore and a promise to regrade the land upon conclusion of the lease. Assume the buyer-lessee breaches the agreement by failing to regrade the land. Courts have divided (bitterly) over the proper damages default.[179] Should the lessor be entitled to cost of performance damages when the economic loss from breach (as measured by the diminution in market value) is substantially less than the cost of regrading? As in the lost volume sales debate, the focus on expectation damages as a majoritarian default is misplaced whenever insurance or risk management are contracting goals. As the parties to these grading contracts trade off the various insurance and information factors identified in Part II, they choose among a wide range of option prices and, if the termination is a breach provision, an equally wide range of damages measures. Only occasionally will the option price reflect the amount of compensation for the lessor's loss.[180]

3. *Is the Damages Default Irrelevant?* — In this Article, we have emphasized that optimal termination rights, including damages, are highly context dependent. This is a familiar result to scholars who evaluate the ability of contract remedies to achieve both ex ante investment efficiency and ex post trade efficiency. We believe that risk management is a critical goal of contracts that has been neglected in the analysis of thin-market

---

§ 2-718(2). Barry Adler and Alan Schwartz argue that section 2-718(2) prevents sellers with market power from contracting for nonrefundable deposits in lost volume contracts where they might otherwise desire to price discriminate. See Adler & Schwartz, supra note 156, at 10–11, 17 (explaining that "[a] legally unconstrained seller will require a down payment that is less than its full expectation and charge a transaction price that is above its cost").

177. The offset for lost volume damages is implicitly invited by section 2-718(3). See U.C.C. § 2-718(3).

178. See, e.g., R.E. Davis Chem. Corp. v. Diasonics, Inc., 826 F.2d 678, 683–84 (7th Cir. 1987); Rodriguez v. Learjet, Inc., 946 P.2d 1010, 1015 (Kan. Ct. App. 1997).

179. Compare Am. Standard, Inc. v. Schectman, 439 N.Y.S.2d 529, 533 (App. Div. 1981) (holding that damages should be calculated at cost of performance even if disproportionate to diminution in value), with Peevyhouse v. Garland Coal & Mining Co., 382 P.2d 109, 114 (Okla. 1962) (holding that damages recoverable were limited to diminution in value of premises rather than cost of performance).

180. The better solution, as we argue below in Part III.D, is for courts to specifically enforce the verifiable terms of the contract as an inducement to parties for whom termination options are efficient to bargain explicitly over the option price.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

contracts. When it is brought into the picture, the heterogeneity of optimal damages is all the more striking and it becomes clear that the expectation damages measure enjoys no privileged status as a majoritarian default rule. It is tempting to argue, therefore, that the only necessary reform is to repeal the penalty rule and specifically enforce all liquidated damages clauses and other termination provisions. This would give parties the freedom to contract away from the damages default. If any given default measure of damages were no more likely to be selected by contracting parties than the next one, the choice among default damages would be irrelevant. Moreover, a default damages measure that conforms to preferences of even a plurality of parties would at least save the costs of contracting in those cases. Expectation damages might be the most plausible candidate on this ground because of its historical and doctrinal salience.

The difficulty, however, is that any damages default imposes costs on all those parties who would prefer to contract out. Moreover, the gain of having a rule appropriate for a plurality of parties yields a lower social gain than one satisfying a majority. Against this limited gain, we must set out the cost to the majority who must contract around the default rather than draft their contract on a blank slate. Courts tend to regard state-created defaults as presumptively fair or efficient, and this institutional bias raises the cost of contracting out.[181] This is particularly true in the case of expectation damages because contract doctrine explicitly trumpets the compensatory purpose of damages and the correspondingly acceptable rationale for liquidated damages—the difficulty of verifying the promisee's loss. The historical and doctrinal salience of expectation damages that makes it an attractive plurality default is likely to cause it to be stickier than most other default provisions.

The stickiness of a compensation default presents an additional impediment to efficient contracting. Lawyers draft termination options for their clients against the background of the expectation damages default that they have internalized since learning the doctrine in law school. This default rule is likely to have the same salience for them as for the courts. Thus, a decision simply to abandon the penalty rule while retain-

181. See, e.g., Davis v. Small Bus. Inv. Co., 535 S.W.2d 740, 744 (Tex. Civ. App. 1976) (concluding that contractual provision purporting to allocate to debtor burden of "all" expenses incurred in preserving collateral is not "agreement otherwise" sufficient to opt out of Texas equivalent of UCC section 9-207(2)(a)); Caudle v. Sherrard Motor Co., 525 S.W.2d 238, 240 (Tex. Civ. App. 1975) (holding that parties must use clear and unequivocal language to shift liability for risk of loss from seller to buyer for period before buyer receives merchandise from seller); see also Hayward v. Postma, 188 N.W.2d 31, 33 (Mich. Ct. App. 1971) (same). Moreover, courts may be reluctant to give the express language of the contract a meaning that conflicts with the relevant default. See Nanakuli Paving & Rock Co. v. Shell Oil Co., 664 F.2d 772, 780 (9th Cir. 1981) (holding that even express price term is insufficient to trump default rule if interpretation of contract does not "totally negate" express term). For discussion, see Goetz & Scott, Limits of Expanded Choice, supra note 161, at 290–91.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

ing the expectation damages default is unlikely to neutralize the doctrinal bias in favor of compensation. Retaining the expectation damages default perpetuates the burden on the many contracting parties who would be better served by adopting different termination provisions.[182]

We believe that any particular damages default, even if divorced from compensation, suffers from stickiness in the courts and at the drafting table. Suppose the default provided for damages equal to ten percent of the contract price. Given the salience of any default, it would lead courts and lawyers to anchor on the default and to resist significant departures from it. As with the expectation damages default, the consequent obstacle to contracting is unlikely to be justified by the relatively few cases in which the default is optimal. In the next section, therefore, we urge a more radical shift in the default to specific performance, as a clear signal to parties that their contract embeds no options or termination rights unless they provide for them explicitly. In enforcing explicit option prices, the courts then will not be held back by the anchor of a damages default.

D. *Alternatives to Damages Defaults*

As indicated earlier, our analysis of embedded options reinforces the case for specific enforcement of liquidated damages and other termination provisions. Absent an explicit agreement over the option price, the courts must have some rule by default to resolve the contract dispute. We have expressed concerns about the enterprise of selecting an alternative default that is anchored on some measure of damages. A preferable approach is to establish instead a bargain-forcing default that not only maximizes the freedom to contract for termination rights but indeed also encourages the parties to bargain explicitly over these provisions. As discussed in the previous section, any damages default runs the risk of anchoring judicial interpretation and legal drafting at the default measure. Therefore, we explore below two alternative bargain-forcing approaches: 1) specific enforcement of exchanges in contracts that lack options to terminate;[183] and 2) non-enforcement of the exchange against the promisor (i.e., where the buyer enjoys a free option to terminate).

1. *A Specific Performance Default in Commercial Contracts.* — In commercial contracts, we propose that specific performance be adopted as the

182. The same criticism may be leveled at the regulation of termination rights under the requirement of good faith. See Goldberg, Discretion, supra note 1, at 347–48 (questioning why good faith is interpreted to result in inflexible default rule that is difficult to contract around).

183. The case for specific performance is best understood by visualizing the termination option as creating a contract plus a put. Where the parties fail to negotiate the exercise price of the put, the "no option" solution would be logically compelled and the buyer would be specifically required to pay the full purchase price for the goods or services.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

general default rule.[184] To be sure, other scholars have argued in favor of a specific performance default, particularly in circumstances where the parties can renegotiate their bargain.[185] And we are not alone in suggesting that courts should specifically enforce any express agreement as to liquidated damages or termination rights.[186] For our purposes, however, it is particularly significant that specific performance also means that the court will not read any termination right into the contract if the parties are not explicit in this respect. In these cases, the court will specifically enforce the exchange alone.[187] We would expect that, in equilib-

184. Granting parties the right to contract for specific performance (other than for personal service contracts) would return the common law to its historical roots described in Part I.B. The device of the penal bond was a means by which parties through the eighteenth century could, in effect, contract for specific performance. The demise of the penal bond therefore eliminated parties' ability to contract for specific performance.

Specific performance appears to be routinely available as the default remedy in most civil law regimes. See Barry Nicholas, The French Law of Contract 211–20 (2d ed. 1992) (discussing specific performance under French law); Guenter H. Treitel, Remedies for Breach of Contract: A Comparative Account 51 (1988) (discussing specific performance under German law); Deborah E. Townsend, The Foreign Economic Contract Law of the People's Republic of China: A New Approach to Remedies, 24 Stan. J. Int'l L. 479, 485 (1988) (discussing necessity of specific performance as a remedy in a planned socialist economy); see also United Nations Convention on Contracts for the International Sale of Goods, Apr. 11, 1980, art. 46(1), S. Treaty Doc. No. 98-9 (1986), 1489 U.N.T.S. 3, 67 ("The buyer may require performance by the seller . . . unless the buyer has resorted to a remedy which is inconsistent with this requirement."). For a discussion of specific performance under the CISG, see Steven Walt, For Specific Performance Under the United Nations Sales Convention, 26 Tex. Int'l L.J. 211, 213–24 (1991).

185. For a persuasive argument in favor of a default remedy of specific performance, see Alan Schwartz, The Case for Specific Performance, 89 Yale L.J. 271 (1979).

186. E.g., Edlin & Schwartz, supra note 103, at 53–54; see also, e.g., Adler & Schwartz, supra note 156, at 27 (arguing that current liquidated damages rules "encourage frivolous law suits" because they "ban a practice—exacting supracompensatory payments for breach—that does not exist").

187. In the case where the buyer is the promisor, a specific enforcement default is straightforward. It authorizes an action by the seller for the price of the contracted goods or services. See, e.g., U.C.C. § 2-709 (2003) (allowing seller to recover price of accepted goods, lost or damaged conforming goods, and goods "identified to the contract" that "seller is unable after reasonable effort to resell"). The issue of defective or inadequate performance by the buyer is not relevant: Either the buyer has paid the price or he will be ordered to do so. On the other hand, where the promisor is the seller, the proposed default is somewhat more complicated. Specific performance would be available to order the seller to perform an executory promise (other than for personal service contracts). In addition, a buyer would have recourse to existing default rules governing implied and express warranties to recover money damages whenever the seller's performance failed to meet the contract requirements. In such a case, the buyer is entitled to damages based on the difference between the actual value of the goods or services and their value as warranted in the contract. A specific performance default will not eliminate the need to calculate damages in these cases of nonconforming performance. Thus, for example, UCC section 2-714 would still be applicable. Under this provision, the buyer who takes possession of nonconforming goods may recover damages for any nonconformity in the seller's performance. U.C.C. § 2-714(1). Under UCC section 2-714(2), "[t]he measure of damages for breach of warranty is the difference between the value of the goods accepted and the value they would have had if they had been as warranted." § 2-714(2). Our

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

rium, parties would respond to the specific performance default by explicitly setting out termination rights in their agreements, particularly given that they will be free from the bias generated by an expectation damages default.

The specific performance approach is well suited for contracts between commercial parties because it closely parallels thick-market contracting. As we observed earlier, thick-market contracts are clearly segmented into forward or future contracts on the one hand and option contracts on the other. If a contract does not specify an option, none will be implied. In this way, a specific performance decree functions to enforce the allocation of market risks just as the award of market damages does, but without the need to verify market prices to a court.[188]

2. *A Bargain-Forcing Rule for Consumer Buyers.* — A specific performance default may not be appropriate in consumer transactions, however. Especially in the case of standardized goods, for which a merchant seller has not made a specific investment, a default that requires the consumer buyer to pay the full purchase price and accept delivery of unwanted goods would serve no efficiency purposes and could work unnecessary hardship on consumer buyers. Merchants also are more likely to be superior risk bearers because of their ability to spread risk among their customers and to control the quality of their products. As a bargain-forcing rule, specific performance is ineffective against consumers who typically do not negotiate the terms of their contracts.

Merchants typically draft consumer contracts, and the legal rule might better focus the incentive to provide explicitly for termination options on them. For other sellers, such as retailers of electronic goods or merchants who rely on direct mail sales, the free-option default would merely replicate their existing terms of sale. We propose, therefore, that absent a specific agreement providing for nonrefundable deposits, liquidated damages, or other termination rights, the default rule should provide consumer buyers free options: the right to walk away from the executory exchange or to hold the merchant seller to the deal. The default

---

proposal will reduce judicial measurement costs in resolving such disputes over executory contracts.

188. As noted in Part III.B, we endorse market damages in thick contracts. Despite the high degree of heterogeneity among thick-market contracts, the market damages default serves the principal contracting goal in these markets—namely, to allocate market risk. Indeed, the designation "market damages" may be a misnomer because, in essence, the default specifically enforces the parties' contract. One option might be to extend the market damages default that applies in thick markets to the thin-market context as well. This proposal has two advantages. First, it eliminates the inevitable boundary problems of distinguishing between thin- and thick-market transactions. Second, market damages have the attractive feature of being a default rule rather than a default standard. See supra notes 163–164 (explaining disadvantages of standards). But this suggestion flounders on the difficulty of verifying the relevant "market price" in transactions in which the market is thin, by definition. The inability to verify market prices would produce the very same litigation cost problems that commended the proposal in the first place.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

would motivate those sellers who prefer to offer buyers a priced option or no option at all to contract out of the default in their standard forms or, in the case of retail sellers, by posting appropriate notification at the point of sale.

This forcing rule is analogous to the statute of frauds in section 2-201 of the UCC that limits recovery to the quantity of goods stated in the contract.[189] If the parties fail to provide for quantity, the contract is effectively unenforceable. As Ian Ayres and Robert Gertner have argued, this forcing rule motivates parties to contract explicitly for a specified quantity.[190] The same argument applies to the price of the embedded option, a price that often functions as damages for breach. If the parties do not provide explicitly for the option price, then the court should not enforce the option. As indicated above, in the commercial case we propose that courts specifically enforce the exchange but not the option.[191] In either case, therefore, the law should be no more worried about the scenario where the parties fail to specify damages (the option price) than it is worried about contracting parties (off the equilibrium path) who fail to specify a quantity term.

To be sure, the free-option default for consumers does require courts to craft a set of associated rules governing the terms and duration of the consumer buyer's call option. But a useful template already exists in the default rules governing a buyer's right to reject defective goods under the UCC. Section 2-602 provides that a rejecting buyer must exercise the right of rejection within a reasonable time, cannot thereafter exercise ownership inconsistent with the seller's rights, and must hold the rejected goods for the seller with reasonable care.[192] Applying a similar default standard to the consumer buyer exercising her termination option would protect both the buyer's option and the seller's interest in the underlying goods or services.

The indefiniteness doctrine of contract law offers an appropriate doctrinal home for a default that would bar contract enforcement against the consumer buyer.[193] Courts declare executory agreements void for indefiniteness if the court is unable to determine the occurrence of

189. U.C.C. § 2-201(1).

190. Ian Ayres & Robert Gertner, Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules, 99 Yale L.J. 87, 95–97 (1989).

191. See supra Part III.D.1.

192. U.C.C. § 2-602(1) to (2). In turn, a failure to reject in time or a wrongful exercise of ownership inconsistent with the seller's interests constitutes an acceptance under section 2-606. Id. § 2-606(1)(b) to (c).

193. Conventional wisdom holds that modern courts will enforce indefinite contracts by filling contractual gaps with general standards of reasonableness and good faith. See, e.g., 1 Corbin, supra note 18, § 95, at 400. This conventional wisdom is misleading, however. A systematic review of the case law shows that American courts continue to dismiss claims for breach of contract on the grounds of indefiniteness, often without granting any relief to the disappointed promisee. See Scott, Self-Enforcing Indefinite Agreements, supra note 160, at 1651.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

breach and provide an appropriate remedy.[194] Unless the parties price a consumer buyer's call option, courts would simply lack "a reasonably certain basis for giving an appropriate remedy" to a merchant seller.[195] The free-option consumer default is similar to the indefiniteness rule that courts apply when parties fail to provide a contract price for services for which there is no established market price.[196] The bargain-forcing default thus uses the threat of non-enforceability to encourage merchant sellers to specify the terms of any embedded option that they choose to write in consumer contracts. Viewed through this lens, a free-option default in thin or hybrid consumer markets would require merchants to address termination and breach as unified issues.

CONCLUSION

Many consumer and commercial contracts contain explicitly negotiated termination provisions. We have analyzed these provisions as embedded options that serve a valuable risk management function. We have focused attention on the buyer's right to terminate as creating a call option and providing insurance against decreases in the buyer's valuation of the exchange. Parties in different circumstances may agree to a wide range of option-price and exercise-price combinations. This perspective allows us to analyze contract damages as embedded call options that serve similar insurance or risk management purposes. Given that contract damages are option prices, the inquiry dictated by the penalty rule—whether stipulated damages reflect the actual or expected loss to the seller when the buyer walks away from the exchange—is wholly inapt.

---

194. See Restatement (Second) of Contracts § 33(2) & cmt. a (1981) ("If the essential terms are so uncertain that there is no basis for deciding whether the agreement has been kept or broken, there is no contract."); U.C.C. § 2-204(3) ("[A] contract for sale does not fail for indefiniteness if the parties have intended to make a contract and there is a reasonably certain basis for giving an appropriate remedy.").

195. See U.C.C. § 2-204(3) (endorsing indefiniteness doctrine when courts lack basis for imposing an appropriate remedy). The indefiniteness rule would apply only where the agreement was executory and the buyer declined to pay the contract price and retain the contract goods or services. On the other hand, if the buyer affirmed the agreement and exercised his option to buy the goods or services, then the conduct of the parties would establish the existence of an enforceable agreement binding on the parties. See, e.g., Restatement (Second) of Contracts §§ 4, 30(1); U.C.C. § 2-207.

196. The UCC authorizes courts to provide a reasonable price where the parties have not specified a price but intended to be bound to the agreement despite the lack of specification in the price. § 2-305(1). This provision, however, simply follows the line of cases holding that price terms in sales contracts can be supplied from evidence of market prices. See, e.g., Varney v. Ditmars, 111 N.E. 822, 825 (N.Y. 1916) (positing in dicta that "parties may use the words 'fair and reasonable value' as synonymous with 'market value'"). But where there is no relevant market price, courts continue to decline to enforce such agreements on the grounds of indefiniteness. See, e.g., Smith v. Hammons, 63 S.W.3d 320, 325–26 (Mo. Ct. App. 2002) (affirming trial court's summary judgment decision in favor of defendant where parties did not agree on essential terms of contract, such as plaintiff's share of profits and value of signing bonus). For discussion, see Scott, Self-Enforcing Indefinite Agreements, supra note 160, at 1647–51, 1655–60.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions

The price of the option is a function of its value to the buyer and its cost to the seller. Only in unusual circumstances will that option price reflect the ex post loss suffered by the seller if the option is exercised.

The option price is set at or above the cost of the option to the seller, depending on the exercise price. A court might ask instead whether stipulated damages are a fair price for the option in light of its exercise price. But that would take the courts into the realm of price regulation, a task they appropriately eschew in other contexts.[197] The implicit premise is that prices are extraordinarily heterogeneous and parties have far superior information than the courts have in this regard. We argue not only that contracting parties should be free to set the prices for their embedded options, but also that the law should not provide default damages rules that effectively establish prices for the parties when they fail to do so for themselves. Rather, in commercial transactions, the courts should specifically enforce contract exchanges in order to encourage parties to bargain over termination rights and embedded options. In light of the fact that merchants typically draft contracts in consumer transactions, we propose that they be encouraged to specify termination rights explicitly by enforcing a contract default that would give consumers free options, i.e., the right to terminate without charge. This rule would motivate merchants to specify at the time of contracting the extent and nature of a consumer's right to return.

In many ways, our proposals reflect nostalgia for the contract law of the early nineteenth century, before the compensation principle led the courts to expand the reach of expectation damages to include lost profits and before they began to aggressively strike down penalty liquidated damages. In this light, our project in this Article is to trim back the unruly branches of the common law. One might argue that our position ignores the dramatic economic changes that have occurred over the past two centuries. To the contrary, however, our approach is justified by concomitant changes—the growing heterogeneity of contractual contexts, the increase in the speed with which parties can access and process new information, and the greater appreciation for the benefits and tools of risk management.

197. Note that while antitrust law polices markets in which prices are determined, there is otherwise no consistently broad attempt by the law to regulate prices.

This content downloaded from 128.59.161.126 on Fri, 18 Dec 2015 19:30:11 UTC
All use subject to JSTOR Terms and Conditions